

BULLOCK

Fitting Curves to Observations.

Graduate School

A. M.

1906

UNIVERSITY OF ILLINOIS
LIBRARY

Class
1906

Book
B876

Volume

Je 06-10M

Return this book on or before the
Latest Date stamped below. A
charge is made on all overdue
books.

University of Illinois Library

UN

L161—H41

78
147

FITTING CURVES TO OBSERVATIONS

BY

JESSIE JANE BULLOCK, A. B. '00

THESIS FOR THE DEGREE OF MASTER OF ARTS
IN THE GRADUATE SCHOOL

IN THE
UNIVERSITY OF ILLINOIS
PRESENTED JUNE 1906



Digitized by the Internet Archive
in 2013

<http://archive.org/details/fittingcurvestoo00bull>

UNIVERSITY OF ILLINOIS

June 2

190 6

THIS IS TO CERTIFY THAT THE THESIS PREPARED UNDER MY SUPERVISION ^{*}BY

Jessie Jane Bullock,

ENTITLED Fitting Curves to Observations

IS APPROVED BY ME AS FULFILLING THIS PART OF THE REQUIREMENTS FOR THE DEGREE

OF Master of Arts

E. J. Gourneau

HEAD OF DEPARTMENT OF Mathematics.

* *With the assistance of Prof. W. L. Rietz.*

INTRODUCTION.

It is the purpose of this thesis to study one of the common mathematical problems of the physicist, astronomer, engineer and statistician, namely to derive the equation of a curve which represents graphically certain observations or experimental data. The point of view differs radically from that of purely theoretical treatment in which one finds that thru any n points arbitrarily chosen he may pass a curve of any form whatever provided its equation contains n constants. The problem here is:- Given n points, to derive with a minimum of labor an equation of the simplest form possible with as few constants as possible, whose locus shall fit these points nearly enough for the purpose for which it is to be used.

The common and almost universally accepted method of dealing with this problem is the method of least squares. This method has the advantage that, given a set of data and an equation containing certain undetermined constants, it so determines these constants as to make the sum of the squares of the residuals a minimum. It has the disadvantage of being impracticable in many cases, and even in those cases where it

is most applicable very laborious, and hence if another method can be found which will give results as accurate as one's data and do it with less labor it is for that set of data a better method. Moreover the method of least squares becomes in some cases where the undetermined constants are not linearly involved hopelessly cumbersome by leading to equations which can not be solved. It is one thing to say that n equations can be solved for n unknowns and another and quite different thing to solve them, as for instance in many exponential and transcendental equations. Probably after the common parabolic form

$$y = a_0 + a_1x + a_2x^2 + \dots + a_nx^n$$

there are few forms more commonly useful than the sine curve and the form with undetermined exponents. But if the period of a sine curve or the unknown exponent is one of two or more constants to be determined the method of least squares fails. Hence there is certainly need for some systematic method of fitting curves of these forms. Moreover the relative ease with which the constants in the common parabolic form may be determined by the method of least squares has led to its indiscriminate use in cases where some other form of function might be more advantageously assumed. Certain cases where other curves are useful and methods of fitting these to data will be presented. While there is no pretension to ex-

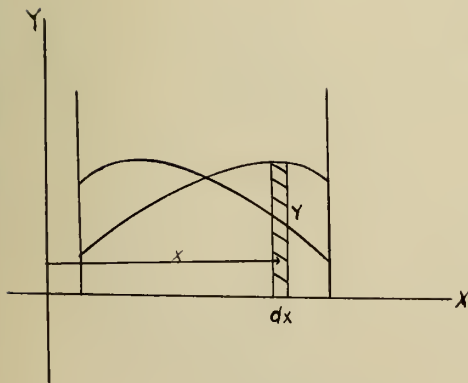
haustive treatment, it is hoped that some of these problems and principles may be suggestive and useful.



THE METHOD OF MOMENTS.

The method of moments as a method of curve fitting is due to Prof. Karl Pearson and has been published by him in the Philosophical Proceedings of the Royal Society, in *Biometrika* and elsewhere.

Considered a priori, in an intuitional rather than ⁱⁿ a mathematical way, there would seem to be some reasonableness to the method. There are given n points. If they conform to some law capable of expression in an equation, the problem is to find that equation. If the equation is to give a perfect fit, then ^{the} theoretical curve and the experimental curve together with certain fixed ordinates at $x = a$ and $x = b$ and with the x -axis bound the same area.



But an indefinitely large number of curves can be drawn which will give the same $\surd$ area under the curve between the same limits as in the accompanying figure. If the first moments of the theoretical and observational curves are equal, then the sums of the products obtained by

multiplying each element of area by its distance from an arbitrarily chosen axis, that is, $\int xy \, dx$ in the two curves are equal. This will give a condition as to the distribution of the area along the x-axis. If second moments, that is, $\int x^2 y \, dx$ are equal, there is a second condition on this distribution of area along the x-axis. If in this way as many conditions be put upon the theoretical curve as it contains undetermined constants it might seem that a method of curve-fitting were here set up. It remains, however, to prove mathematically that this method of procedure actually will fit a curve to observations, and to show what degree of accuracy may be expected under different conditions as to data, and form of function assumed.

Given a series of observations of a variable y , corresponding to a series of values of a second variable x , to show that the method of moments furnishes a good method of fitting to this data a theoretical curve

$$y = f(x, c_1, c_2, \dots, c_n)$$

where the c 's are arbitrary constants. Let it be assumed that the function can be expanded by Taylor's theorem and converges fairly rapidly. Let the expansion be

$$\begin{aligned} y &= f(0) + f'(0)x + \frac{f''(0)x^2}{2!} + \frac{f'''(0)x^3}{3!} - - - \\ &= a_0 + a_1x + \frac{a_2x^2}{2!} + \frac{a_3x^3}{3!} + - - - \end{aligned}$$

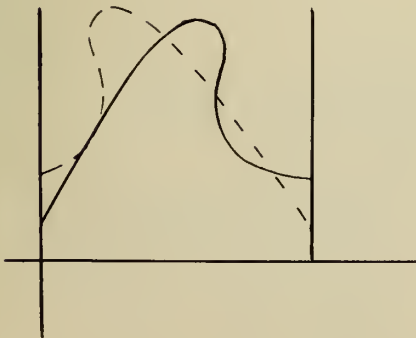


Each a will be a function of the c 's of the original equation. Hence when each of the first n a 's ^{is} expressed in terms of the n c 's, there arise n equations which may be solved for the c 's in terms of the first n a 's. Then the succeeding a 's, i.e., those after a_{n-1} , may be expressed first in terms of the c 's, then by substitution in terms of the first n a 's.

Then

$$y = a_0 + a_1x + \frac{a_2x^2}{2!} + \dots + \frac{a_{n-1}x^{n-1}}{(n-1)!} \\ + f^{(n)}(a_0, a_1, \dots, a_{n-1}) \frac{x^n}{n!} + f^{(n+1)}(a_0, \dots, a_{n-1}) \frac{x^{n+1}}{(n+1)!} \\ + \dots$$

Let y' be the observed, y the theoretical value. Then $y - y'$ is the difference between the observed and theoretical curves at a point corresponding to a given x . If $\int (y - y')^2 dx$ be made a minimum, $y - y'$ will be made small. It would not be of much service to make $\int (y - y') dx$ a minimum for even if $\int (y - y') dx = 0$, it only means that the areas under the curves are equal, and there would be nothing to exclude curves like those in figure. $y - y'$ is a part of the time positive, the other part negative and the two quantities exactly cancel each other, although the curves are entirely different. But $(y - y')^2$ is necessarily



positive and since $\int (y - y')^2 dx$ involves the summing of all the necessarily positive products of $(y - y')^2$ and dx , making $\int (y - y')^2 dx$ a minimum is bound to bring the theoretical and experimental curves as near together as the given data and the assumed form of equation will permit.

To make $\int (y - y')^2 dx$ a minimum, its first derivative must be zero. This makes

$$\int (y - y') \delta y dx = 0$$

for y is the variable and differentiation under the integral sign can be performed by ordinary method of differentiation if the limits of integration are independent of the quantity with respect to which the differentiation is performed. δy is the variation in y due to the variations in the coefficients $a_0, a_1, \dots, a_{n-1}$.

$$\begin{aligned} \delta y &= \delta a_0 + \delta a_1 x + \delta a_2 \frac{x^2}{2} + \dots + \delta a_{n+1} \frac{x^{n+1}}{(n+1)} \\ &+ \left(\frac{\partial f^{(n)}}{\partial a_0} \delta a_0 + \frac{\partial f^{(n)}}{\partial a_1} \delta a_1 + \frac{\partial f^{(n)}}{\partial a_2} \delta a_2 + \dots + \frac{\partial f^{(n)}}{\partial a_{n-1}} \delta a_{n-1} \right) \frac{x^n}{n} \\ &+ \left(\frac{\partial f^{(n+1)}}{\partial a_0} \delta a_0 + \frac{\partial f^{(n+1)}}{\partial a_1} \delta a_1 + \frac{\partial f^{(n+1)}}{\partial a_2} \delta a_2 + \dots + \frac{\partial f^{(n+1)}}{\partial a_{n-1}} \delta a_{n-1} \right) \frac{x^{n+1}}{(n+1)} \\ &+ \dots \\ &= \delta a_0 \left(1 + \frac{\partial f^{(n)}}{\partial a_0} \frac{x^n}{n} + \frac{\partial f^{(n+1)}}{\partial a_0} \frac{x^{n+1}}{(n+1)} + \dots \right) \\ &+ \delta a_1 \left(x + \frac{\partial f^{(n)}}{\partial a_1} \frac{x^n}{n} + \frac{\partial f^{(n+1)}}{\partial a_1} \frac{x^{n+1}}{(n+1)} + \dots \right) \end{aligned}$$

$$\begin{aligned}
& + \delta a_2 \left(\frac{x^2}{2} + \frac{\partial f^{(n)}}{\partial a_2} \cdot \frac{x^n}{n} + \frac{\partial f^{(n+1)}}{\partial a_2} \cdot \frac{x^{n+1}}{n+1} + \dots \right) \\
& + \dots \\
& + \delta a_{n-1} \left(\frac{x^{n-1}}{n-1} + \frac{\partial f^{(n)}}{\partial a_{n-1}} \cdot \frac{x^n}{n} + \frac{\partial f^{(n+1)}}{\partial a_{n-1}} \cdot \frac{x^{n+1}}{n+1} + \dots \right)
\end{aligned}$$

Let $R = f^{(n)}(\theta x) \frac{x^n}{n!}$ be the remainder after n terms. Then

$0 < \theta < 1$. Also the coefficient of any one of the a 's after the first term is the partial derivative of R with respect to the a to which that particular a belongs.

$$\begin{aligned}
\therefore \delta y &= \delta a_0 \left(1 + \frac{\partial R}{\partial a_0} \right) + \delta a_1 \left(x + \frac{\partial R}{\partial a_1} \right) + \delta a_2 \left(\frac{x^2}{2} + \frac{\partial R}{\partial a_2} \right) \\
&+ \dots + \delta a_{n-1} \left(\frac{x^{n-1}}{n-1} + \frac{\partial R}{\partial a_{n-1}} \right)
\end{aligned}$$

Substituting the value of δy in

$$\int (y - y') \delta x \, dx = 0$$

and collecting coefficients

$$\begin{aligned}
& \left\{ \int (y - y') \left(1 + \frac{\partial R}{\partial a_0} \right) dx \right\} \delta a_0 + \left\{ \int (y - y') \left(x + \frac{\partial R}{\partial a_1} \right) dx \right\} \delta a_1 \\
& + \left\{ \int (y - y') \left(\frac{x^2}{2} + \frac{\partial R}{\partial a_2} \right) dx \right\} \delta a_2 + \dots + \left\{ \int (y - y') \left(\frac{x^{n-1}}{n-1} + \frac{\partial R}{\partial a_{n-1}} \right) dx \right\} \delta a_{n-1} = 0
\end{aligned}$$

Since $a_0, a_1, \dots, a_{n-1}$ are independent of each other, i.e., since we may vary the coefficients independently of each other, this equation must be an identity and each coefficient is equal to zero.

$$\therefore \int (y-y') \left(1 + \frac{\partial R}{\partial a_0} \right) dx = 0$$

$$\int (y-y') \left(x + \frac{\partial R}{\partial a_1} \right) dx = 0$$

$$\int (y-y') \left(\frac{x^2}{2} + \frac{\partial R}{\partial a_2} \right) dx = 0$$

I

— — — .

$$\int (y-y') \left(\frac{x^{n-1}}{n-1} + \frac{\partial R}{\partial a_{n-1}} \right) dx = 0$$

Let A be the area, AM_1 , AM_2 , - - - AM_{n-1} the first $(n-1)$ moments of the theoretical curve, A' , the area and $A'M'_1$, $A'M'_2$, - - - $A'M'_{n-1}$ like moments of the observation curve about the y -axis which of course may be placed at any point where it may be found convenient.

Performing the indicated integrations so far as possible, and clearing of fractions, gives

$$A = A' - \int (y-y') \frac{\partial R}{\partial a_0} dx$$

$$AM_1 = A'M'_1 - \int (y-y') \frac{\partial R}{\partial a_1} dx$$

$$AM_2 = A'M'_2 - \frac{1}{2} \int (y-y') \frac{\partial R}{\partial a_2} dx$$

II

$$AM_3 = A'M'_3 - \frac{1}{3} \int (y-y') \frac{\partial R}{\partial a_3} dx$$

$$AM_4 = A'M'_4 - \frac{1}{4} \int (y-y') \frac{\partial R}{\partial a_4} dx$$

— — — — —

$$AM_{n-1} = A'M'_{n-1} - \frac{1}{n-1} \int (y-y') \frac{\partial R}{\partial a_{n-1}} dx$$

The term involving the integral will be small for:-

1. It involves the $y-y'$ which is small because

$\int (y-y')^2 dx$ has been made a minimum.

2. $R = \frac{x^n}{n!} f^n(\theta x)$ is by hypothesis small if n be not very small, for the series was rapidly convergent.

3. If the first $n-1$ moments of a curve are equal succeeding moments become more and more nearly equal the larger n becomes, for we put more and more conditions on the distribution of the equal total areas along the x -axis. But

$$\int (y-y') \frac{\partial R}{\partial a_s} dx$$

vanishes if the higher moments are equal.

$$\text{For } R = f^{(n)}(0) \frac{x^n}{n!} + f^{(n+1)}(0) \frac{x^{n+1}}{(n+1)!} + \dots$$

$$\begin{aligned} \therefore \int (y-y') \frac{\partial R}{\partial a_s} dx &= \int (y-y') \frac{\partial f^{(n)}(0)}{\partial a_s} \frac{x^n}{n!} dx + \int (y-y') \frac{\partial f^{(n+1)}(0)}{\partial a_s} \frac{x^{n+1}}{(n+1)!} + \dots \\ &= \frac{1}{n!} \frac{\partial f^{(n)}(0)}{\partial a_s} [AM_n - A'M_n] + \frac{1}{(n+1)!} \frac{\partial f^{(n+1)}(0)}{\partial a_s} [AM_{n+1} - A'M_{n+1}] \\ &\quad + \dots \end{aligned}$$

But $A = A'$ and the factors $M_n - M'_n, M_{n+1} - M'_{n+1}$ are small because the higher moments of two curves having their first $n-1$ moments equal are nearly equal.

Neglecting in II the term which contains these three small factors gives

$$A = A'$$

$$AM_1 = A'M'_1$$

$$\text{III} \quad AM_2 = A'M'_2$$

- - - - -

$$AM_{n-1} = A'M'_{n-1}$$

if the expansion by Taylor's Theorem has a finite number of terms, n , as in common parabolic curve $R = 0$, and this is not an approximation but is perfectly accurate. That is, the n constants of the theoretical curve are determined by equating its area and first $n-1$ moments to the area and first $n-1$ moments of the observed curve. This gives n equations to determine the n constants.

Comparing the method with least squares method shows that the fundamental difference lies in the choice of the quantity minimized. In least squares it is the sum of the squares of the distances between the theoretical and observed curves measured along the ordinates of the points that represent the plotted data. By the method of moments the quantity minimized is $\int (y-y')^2 dx$, that is, the sum of the squares of all the distances between the curve measured on every ordinate. For this reason the method of least squares aims at getting a curve that

will strike all the observed points as well as possible, while the method of moments aims at fitting the curve to a hypothetically existing curve which is the expression of the law sought and whose area and moments are the same as those of the observed curve. It will thus tend to smooth out any irregularities in the data and while it may not come so close to some irregularly placed point, may by that very fact better express the law involved in the data, just as it is universally recognized that a smooth curve drawn through a number of points representing experimental data is a better expression of the law than the polygon made by connecting up the points. Theoretically there is no limit to the degree of exactness of fitting any set of points. The curve may be chosen with more and more constants and by equating areas and moments as many equations are formed as there are constants to be determined. Practically the limit is found in the rapid increase of probable error in the higher moments, as the number of dimensions increases.

The three essentials for the use of this method are then:-

a. We must be able to ascertain the moments of the theoretical curve in terms of its constants.

b. We must know how to find area and moments of any system of observations, i.e., quadrature formulas.

c. The expressions for moments of the theoretical curve in terms of its constants must give equations capable of numerical solution.

As to the form of function to be assumed in any case it may be said that for frequency curves Pearson has worked out a set of five types of curves derived from the point binomial and the theory of probability.

(Philosophical Transactions of the Royal Society, Volume 186A pp. 344-414.)

Merely as illustrating and applying one of the simpler forms which Prof. Pearson has used in his unified and systematic treatment of frequency curves in general a problem in school attendance will be fitted with a point binomial. But the paper is in the main concerned with the other class of problems, those belonging to the physicist and engineer rather than the statistician. Here obviously the forms of curve will be very different from those occurring in statistical work. In some cases the form of the curve is dependent upon some known physical law which leaves only certain constants to be determined, as the coefficient of discharge or coefficient of velocity in a nozzle of a fire-hose. Sometimes a mere inspection of the plotted points will suggest a type of function, as for instance a manifest periodicity may give the general effect of a sine curve. But in many cases the assumption of the

270

255

240

225

210

195

180

165

150

135

120

105

90

75

60

45

30

15

0

Altitudes in $\frac{1}{1024}$'s.

$$\left(\frac{1}{2} + \frac{1}{2}\right)^{10}$$

1

2

3

4

5

6

7

8

9

10

11



form of the function will be more or less arbitrary, and sometimes two quite different forms of function appear to give equally good or equally bad fits over a certain rather limited range of observations. A discussion of forms of functions suited to certain data, based on the behavior of successive orders of differences will be given in section at the end of this thesis.

ON FREQUENCY CURVES.

It is a familiar fact in the theory of probability that the k th term of $(1/2 + 1/2)^n = (1/2)^n + n(1/2)^n + \frac{n(n-1)}{2!}(1/2)^n + \dots + (1/2)^n$ represents the probability of throwing $k-1$ heads and $(n-k+1)$ tails in n throws of a coin. Also that for fairly large values of n as from 10 to 20, if these numbers be plotted as ordinates at equal intervals along the axis of abscissae the points lie very close to the normal frequency curve $y = ke^{-x^2/2\sigma^2}$ where σ^2 and k are constants dependent on n . The sum of all the terms is obviously unity, corresponding to the certainty that some one of these events must happen. Hence if rectangles be constructed of unit base, and altitudes equal to the successive terms of the expansion, the area of the k th rectangle represents the probability of the occurrence of $k-1$ heads and $(n-k+1)$ tails in n throws. Thus if $n = 10$ the successive rectangular areas from left to right represent the probability of throwing 0, 1, 2, 3, ..., 10 heads in 10 throws. The curve in dotted lines represents approximately this same area under a continuous curve very like $y = ke^{-x^2/2\sigma^2}$.

This represents a case where the probabilities of the two elementary events namely that any coin fall as a head or as a tail are equal, hence the probabilities of a heads and b tails is just equal to the probability of b heads and a tails and the curve is symmetrical.

If, however, we consider the probability of throwing an ace with a single die it is $1/6$. By the same theory as the preceding the k th term of $(5/6 + 1/6)^n$ represents the probability of throwing $k-1$ aces. This manifestly will give an asymmetrical curve if plotted in the same manner as before. And the type of curve formed by plotting as ordinates $\frac{Ax}{c}$ successive terms of $(p+q)^n$ where $p+q = 1$ and A is the area under the curve, at intervals c along the x -axis is used by Prof. Pearson to fit data of the nature of frequency curves which have limited range in both directions with skewness. A , the area, represents the total number of observations.

Problem- To fit the best point binomial curve to this set of data. Number of children in Boston school by ages.

Age.	Number.	Age.	Number.	Age.	Number.
3- 4	205	9-10	7908	15-16	3800
4- 5	1765	10-11	7805	16-17	2301
5- 6	5101	11-12	7526	17-18	1414
6- 7	7513	12-13	7198	18-19	747
7- 8	7779	13-14	6862	19-	415
8- 9	7939	14-15	5465		

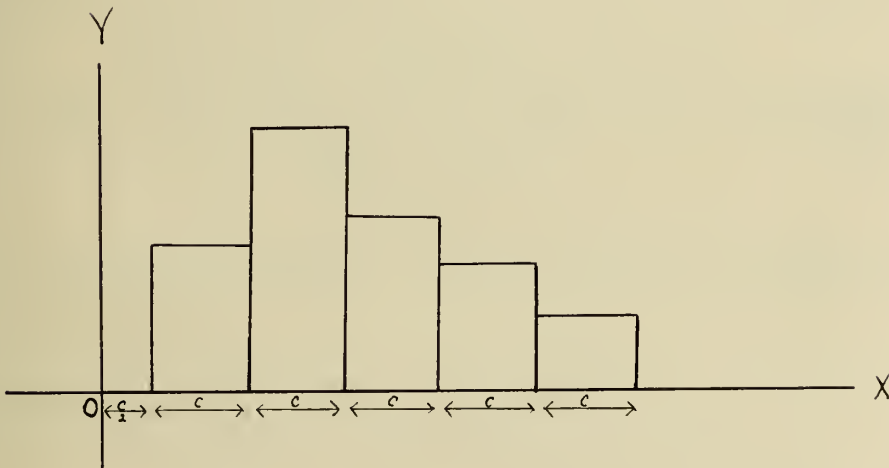
Plotting the data of the school-attendance problem shows asymmetry or skewness hence the applicability of the point binomial curve with $p \neq q$.

Let the figure be considered as a series of rectangles $n+1$ in number of base c , and the altitudes A/c times successive terms of $(p+q)^n$, that is, the r th altitude

$$y_r = \frac{A}{c} \frac{n(n-1) \dots (n-r+2)}{(r-1)!} p^{n-r+1} q^{r-1}$$

The area of any one of the elementary rectangles =

$$cy_r = A \frac{n(n-1) \dots (n-r+2)}{(r-1)!} p^{n-r+1} q^{r-1}$$



Draw a y-axis $1/2 c$ to the left of the initial point of the first rectangle. If then instead of multiplying each element of area by the n th power of its x coordinate and adding the products by $\int x^n y dx$ to get the n th moment we multiply the n th power of the average value of x by the whole area we get a fair approximation for the value of the n th moment.

Using $AM_0, AM_1, AM_2, \dots, AM_n$ to represent the area and first n moments, we get, since the distance is in the r th rectangle is rc ,

$$\begin{aligned} AM_s &= \sum_{r=1}^{r=n+1} [cy_r \times (rc)^s] \\ &= Ac^s \left[p^n + 2^s np^{n-1}q + 3^s \frac{n(n-1)}{12} p^{n-2}q^2 \right. \\ &\quad \left. + \dots + (n+1)^s q^n \right] \end{aligned}$$

for A in the value of cy_r and c^s in $(rc)^s$ is a constant factor common to every term and p^n the remaining factor of the first area is multiplied by the s th power of its value of r , which is unity, $np^{n-1}q$ by s th power of its value of r , and so on throughout the expansion.

Substituting value of $p = 1-q$

$$\begin{aligned} AM_s &= Ac^s \left[(1-q)^n + 2^s nq(1-q)^{n-1} + 3^s \frac{n(n-1)}{12} q^2(1-q)^{n-2} \right. \\ &\quad \left. + \dots + (n-1)^s q^n \right] \end{aligned}$$

Collecting like powers of q and afterward simplifying the coefficients

$$\begin{aligned} AM_s &= Ac^s \left[1 + q(-n + 2^s n) \right. \\ &\quad \left. + q^2 \left(\frac{n(n-1)}{12} - 2^s n(n-1) + 3^s \frac{n(n-1)}{12} \right) \right. \\ &\quad \left. + q^3 \left(-\frac{n(n-1)(n-2)}{12} + 2^s \frac{n(n-1)(n-2)}{12} - 3^s \frac{n(n-1)(n-2)}{12} + 4^s \frac{n(n-1)(n-2)}{12} \right) \right. \\ &\quad \left. + q^4 \left(\frac{n(n-1)(n-2)(n-3)}{12} - 2^s \frac{n(n-1)(n-2)(n-3)}{12} + 3^s \frac{n(n-1)(n-2)(n-3)}{12 \cdot 12} \right. \right. \\ &\quad \left. \left. - 4^s \frac{n(n-1)(n-2)(n-3)}{12} + 5^s \frac{n(n-1)(n-2)(n-3)}{12} \right) + \dots \right] \end{aligned}$$

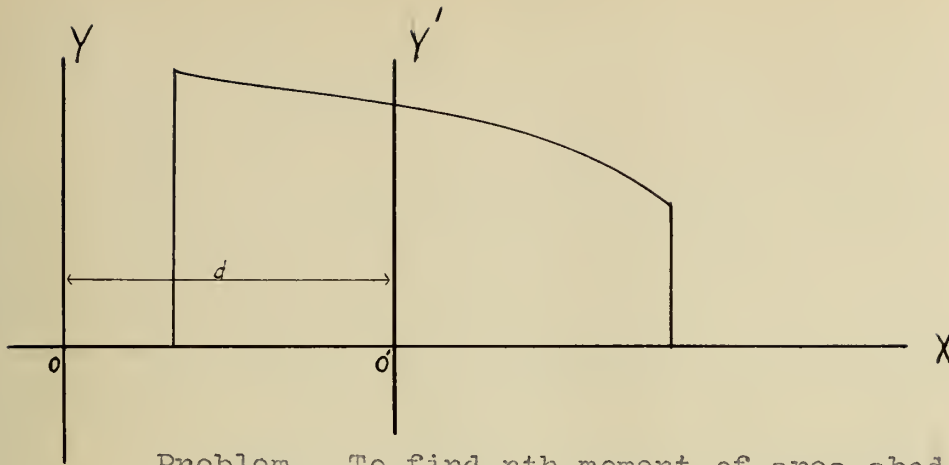
$$\begin{aligned}
&= Ac^s \left[1 + nq(-1+2^s) + \frac{n(n-1)}{2} q^2 (1-2 \cdot 2^s + 3^s) \right. \\
&\quad + \frac{n(n-1)(n-2)}{6} q^3 (-1+3 \cdot 2^s - 3 \cdot 3^s + 4^s) \\
&\quad \left. + \frac{n(n-1)(n-2)(n-3)}{24} q^4 (1-4 \cdot 2^s + 6 \cdot 3^s - 4 \cdot 4^s + 5^s) + \dots \right]
\end{aligned}$$

Using this formula to derive the first four moments necessary to determine the four quantities p , q , n and c of the point binomial,

$$\begin{aligned}
AM_1 &= Ac(1+nq) \\
AM_2 &= Ac^2 \left[1 + 3nq + n(n-1)q^2 \right] \\
AM_3 &= Ac^3 \left[1 + 7nq + 6n(n-1)q^2 + n(n-1)(n-2)q^3 \right] \\
AM_4 &= Ac^4 \left[1 + 15nq + 25n(n-1)q^2 + 10n(n-1)(n-2)q^3 \right. \\
&\quad \left. + n(n-1)(n-2)(n-3)q^4 \right]
\end{aligned} \tag{IV}$$

for the $(s+1)$ st and all succeeding polynomials vanish identically.

But the moments are taken about an axis distant $1/2 c$ from initial point of the first rectangle and c is an unknown. In order there that we may take moments of the observed curve and of the theoretical curve about same axis some new axis common to both must be determined. Since obviously the centroids of the areas must coincide a centroidal axis of moments is convenient. Since this change of axis is often necessary, formulas are here derived for transfer of moments from any axis to any other and then a special and somewhat simplified form for the case when the new axis passes through the centroid.



Problem.- To find n th moment of area abcd about y' -axis is in terms of its moments about the y -axis.

Let AM_n be n th moment about y -axis.

AM'_n be n th moment about y' -axis.

d be n th distance between these axes.

$$\begin{aligned} \text{Then } AM'_n &= \int x'^n y \, dx = \int (x-d)^n y \, dx \\ &= \int x^n y \, dx - dn \int x^{n-1} y \, dx + d^2 \frac{n(n-1)}{2} \int x^{n-2} y \, dx + \dots + d^n \int y \, dx \\ &= AM_n - dn AM_{n-1} + d^2 \frac{n(n-1)}{2} AM_{n-2} + \dots + d^n AM_0 \end{aligned}$$

Dividing out A

$$M'_n = M_n - dn M_{n-1} + d^2 \frac{n(n-1)}{2} M_{n-2} + \dots + d^n$$

since M_0 is always unity and is introduced merely for symmetry of formulas.

If now the y' -axis be centroidal, then by definition the first moment vanishes and

$$M'_1 = 0$$

But

$$M'_1 = M_1 - M_0 d$$

$$0 = M_1 - d$$

$$d = M_1$$

which is ordinary method of getting $\bar{x}$.

$$\begin{aligned} M_2' &= M_2 - 2dM_1 + d^2 \\ &= M_2 - d^2 \end{aligned}$$

V

which is ordinary formula used in mechanics for transferring moments of inertia to centroidal axis.

In general, this equality of $d = M_1$ will shorten the formulas by combining the last two terms, and

$$\begin{aligned} M_3' &= M_3 - 3dM_2 + 2d^3 \\ M_4' &= M_4 - 4dM_3 + 6d^2M_2 - 3d^4 \end{aligned}$$

VI

Applying these formulas to equations IV

$$\begin{aligned} d &= M_1 = c(1+nq) \\ M_1' &= 0 \\ M_2' &= M_2 - d^2 \\ &= c^2[1+3nq+n(n-1)q^2] - c^2[1+2nq+n^2q^2] \\ &= c^2[nq-nq^2] = c^2nq(1-q) = c^2npq \\ M_3' &= c^3[1+7nq+6n(n-1)q^2+n(n-1)(n-2)q^3] \\ &\quad - 3c^3[1+3nq+n(n-1)q^2][1+nq] \\ &\quad + 2c^3[1+nq]^3 \\ &= c^3[1+7nq+(6n^2-6n)q^2+(n^3-3n^2+2n)q^3] \\ &\quad - c^3[3+12nq+(12n^2-3n)q^2+(3n^3-3n^2)q^3] \\ &\quad + c^3[2+6nq+6n^2q^2+2n^3q^3] \\ &= c^3[nq-3nq^2+2nq^3] \\ &= c^3nq[1-3q+2q^2] = c^3nq(1-q)(1-2q) \\ &= c^3npq(p-q) \end{aligned}$$

$$\begin{aligned}
M'_4 &= c^4 [1 + 15nq + 25n(n-1)q^2 + 10n(n-1)(n-2)q^3 + n(n-1)(n-2)(n-3)q^4 \\
&\quad - 4c^4 [1 + 7nq + 6n(n-1)q^2 + n(n-1)(n-2)q^3] [1 + nq] \\
&\quad + 6c^4 [1 + 3nq + n(n-1)q^2] [1 + 2nq + n^2q^2] \\
&\quad - 3c^4 [1 + 4nq + 6n^2q^2 + 4n^3q^3 + n^4q^4] \\
&= c^4 [1 + 15nq + (25n^2 - 25n)q^2 + (10n^3 - 30n^2 + 20n)q^3 \\
&\quad + (n^4 - 6n^3 + 11n^2 - 6n)q^4] \\
&\quad - c^4 [4 + 32nq + (52n^2 - 24n)q^2 + (28n^3 - 36n^2 + 8n)q^3 \\
&\quad + (4n^4 - 12n^3 + 8n^2)q^4] \\
&\quad + c^4 [6 + 30nq + (48n^2 - 6n)q^2 + (30n^3 - 12n^2)q^3 + (6n^4 - 6n^3)q^4] \\
&\quad - c^4 [3 + 12nq + 18n^2q^2 + 12n^3q^3 + 3n^4q^4] \\
&= c^4 [nq + (3n^2 - 7n)q^2 + (-6n^2 + 12n)q^3 + (3n^2 - 6n)q^4] \\
&= c^4 nq [1 + (3n-7)q + (-6n+12)q^2 + (3n-6)q^3] \\
&= c^4 nq(1-q) [1 + (3n-6)q - (3n-6)q^2] \\
&= c^4 npq [1 + 3(n-2)q(1-q)] \\
&= c^4 npq [1 + 3(n-2)pq]
\end{aligned}$$

There are now four equations to determine four unknown elements of the point binomial.

$$p + q = 1$$

$$M'_2 = c^2 npq$$

$$M'_3 = c^3 npq(p-q)$$

$$M'_4 = c^4 npq [1 + 3(n-2)pq]$$

VII

Eliminating c ,

$$\frac{(M_2')^2}{M_4'} = \frac{npq}{1+3(n-2)pq}$$

$$\frac{(M_2')^3}{(M_3')^2} = \frac{npq}{(p-q)^2}$$

Eliminating n ,

$$npq = \frac{(M_2')^3 (p-q)^2}{(M_3')^2}$$

$$n = \frac{(M_2')^3 (p-q)^2}{(M_3')^2 pq}$$

Substituting in (1)

$$\begin{aligned} \frac{(M_2')^2}{M_4'} &= \frac{\frac{(M_2')^3 (p-q)^2}{(M_3')^2}}{1+3\left(\frac{(M_2')^3 (p-q)^2}{(M_3')^2 pq} - 2\right)pq} \\ &= \frac{(M_2')^3 (p-q)^2}{(M_3')^2 + 3(M_2')^3 (p-q)^2 - 6(M_3')^2 pq} \\ &= \frac{(M_2')^3 (p-q)^2}{(M_3')^2 (1-6pq) + 3(M_2')^3 (p-q)^2} \end{aligned}$$

Dividing out $(M_2')^2$ and eliminating p_1

$$\frac{1}{M_4'} = \frac{M_2' (1-2q)^2}{(M_3')^2 (1-6q+6q^2) + 3(M_2')^3 (1-2q)^2}$$

$$(M_3')^2 (1-6q+6q^2) + (3(M_2')^3 - M_2' M_4') (1-4q+4q^2) = 0$$

$$q^2 [4M_2' (3(M_2')^2 - M_4') + 6(M_3')^2] - q [4M_2' (3(M_2')^2 - M_4') + 6(M_3')^2] + M_2' (3(M_2')^2 - M_4') + (M_3')^2 = 0$$

$$q^2 - q + \frac{M_2' (3(M_2')^2 - M_4') + (M_3')^2}{4M_2' (3(M_2')^2 - M_4') + 6(M_3')^2} = 0$$

VIII

where obviously the sum of the roots of the quadratic = 1.

Hence one root may be taken for q and the other for p . Hence

$pq = k$, constant term of the equation and $p-q = \frac{\sqrt{b^2 - 4ac}}{a}$ from theory of quadratic equations

$$p-q = \sqrt{1-4k} = \sqrt{\frac{2(M_3')^2}{4M_2'(3(M_2')^2 - M_4') + 6(M_3')^2}} = \sqrt{\frac{(M_3')^2}{2M_2'(3(M_2')^2 - M_4') + 3(M_3')^2}}$$

From VII

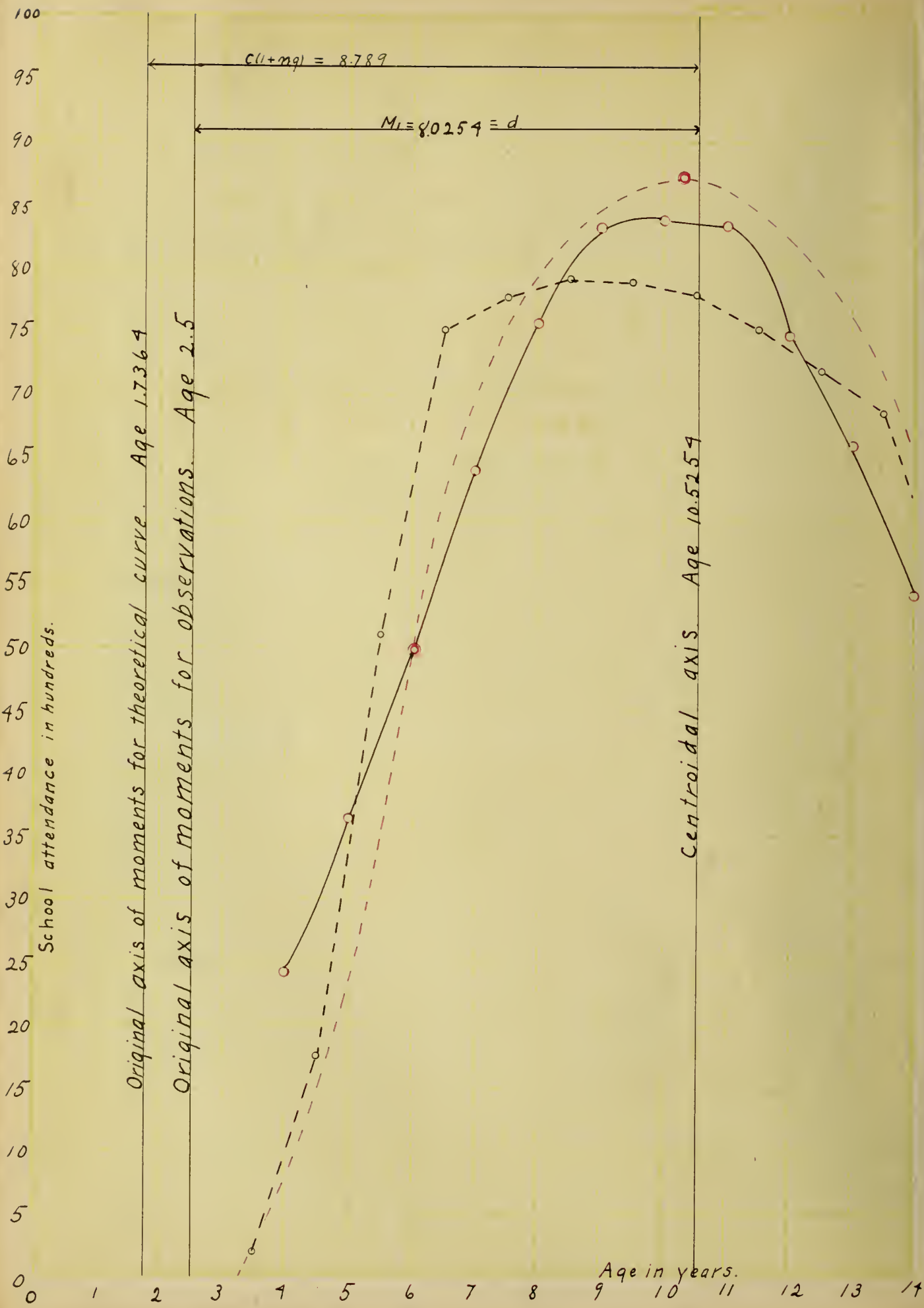
$$\frac{M_3'}{M_2'} = C(p-q)$$

$$C = \frac{M_3'}{M_2'(p-q)} = \frac{\sqrt{2M_2'(3(M_2')^2 - M_4') + 3(M_3')^2}}{M_2'} \quad \text{IX}$$

$$n = \frac{M_2'}{C^2 pq} = \frac{2(M_2')^3}{M_2'(3(M_2')^2 - M_4') + (M_3')^2} \quad \text{X}$$

The area and moments of the observation curve are computed on the same assumption as are those of the theoretical curve.

The data is represented by the figure of plate I. Each space along the x-axis represents a year of age and erecting on each a rectangle whose ordinate is the number of children of that age in the Boston schools, the area = $xy = y$ = number pupils of that age in the schools. Then the sum of these y's gives the total area under the curve. If the area be considered concentrated along the mid-vertical and origin taken at the age $2\frac{1}{2}$ years, then the sum of the products xy gives first moment, the sum of the products x^2y , the second, and so on up to the number needed for the problem, in this case the first four.



Attendance by ages in Boston Public Schools.

--o-- Observations.

-⊙- Point Binomial.

Generalized Probability Curve.

—●— $y = 109^{-1}(-129.941)(x+36.966)^{85.391}e^{-2.31x}$





x	y	xy	x ² y	x ³ y	x ⁴ y
1	205	205	205	205	205
2	1765	3530	7060	14120	28240
3	5101	15303	45909	137727	413181
4	7513	30052	120208	480832	1923328
5	7779	38895	194475	972375	4861875
6	7939	47634	285804	1714824	10288944
7	7908	55356	387492	2712444	18967108
8	7805	62440	499520	3996160	31969280
9	7526	67734	609606	5486454	49378086
10	7198	71980	719800	7198000	71980000
11	6862	75482	830302	9133322	100466542
12	5465	65580	786960	9443520	113322240
13	3800	49400	642200	8348600	108531800
14	2301	32214	450996	6313944	88395216
15	1414	21210	316150	4772250	71583750
16	747	11952	191232	3059712	48955392
17	<u>415</u>	<u>7055</u>	<u>119935</u>	<u>2038895</u>	<u>34661215</u>
		81743	656022	6209854	65823384
					755746402

$$AM_0 = A = 81743$$

$$AM_1 = 656022$$

$$AM_2 = 6209854$$

$$AM_3 = 65823384$$

$$AM_4 = 755746402$$

The M's are found by division and by definition of centroid,

M_1 = distance of the centroid from the assumed axis. Trans-

ferring moments to a centroidal axis by the formulas VI,

$$M'_2 = M_2 - d^2 = 11.5610$$

$$M'_3 = M_3 - 3M_2d + 2d^3 = 10.0113$$

$$M'_4 = M_4 - 4M_3d + 6M_2d^2 - 3d^4 = 308.0799$$

Substituting in equation (VIII), the constant term = $pq = .239766$ whence

$$p = .601$$

$$q = .399$$

Substituting for other values in formulas IX and X,

$$c = 4.287$$

$$n = 2.632$$

All the elements of the point binomial are now known and the terms are,

$$\begin{aligned} p^n &= (.601)^{2.632} \\ n p^{n-1} q &= 2.632 (.601)^{1.632} \times .399 \\ \frac{n(n-1)}{1 \cdot 2} p^{n-2} q^2 &= \frac{2.632 \times 1.632}{2} (.601)^{.632} \times (.399)^2 \\ \frac{n(n-1)(n-2)}{1 \cdot 2 \cdot 3} p^{n-3} q^3 &= \frac{2.632 \times 1.632 \times .632}{1 \cdot 2 \cdot 3} (.601)^{-.368} \times (.399)^3 \end{aligned}$$

The next term would contain the negative factor $n-3$ and hence can not be used. Calculating these terms by logarithms gives them in order 4993, 8723, 4726, 661. To determine the position of these terms to the right or left of the centroidal axis, it is remembered that the first assumed axis of the theoretical curve was $1/2 c$ to the left of the mid vertical of the first rectangle. But the axis was moved $c(1+nq)$ to the right. Therefore the mid-vertical of the first rectangle is $c(1+nq) - c = cnq = 4.502$ to the left of the centroidal axis. But this centroidal axis lies at the age $2.5 + 8.025 =$

10.525. Hence the first term is to be plotted at 6.023. The others are at equal intervals $c = 4.287$ and lie at 10.310, 14.597, 18.884. The curve does not in the figure appear to give a very good fit at the top but the area in the four rectangles of the theoretical curve = 81895.

$$\begin{aligned}\text{observed curve} &= \underline{81743} \\ \text{difference} &= 152\end{aligned}$$

which is less than 1/5 %. It is not possible here to compare individual ordinates for the curve is not continuous as a curve of the form $y = f(x)$. A smooth curve can of course be sketched in connecting the computed points as in the figure.

It is desired now to pass from this point binomial curve to an ordinary analytical geometry curve $y = f(x)$. As an illustration of the method the normal probability curve $y = ce^{-x^2/2\sigma^2}$ will be derived from the symmetrical point binomial, to which its resemblance is well known, and then an analogous method will be taken to derive the generalized probability curve from the asymmetrical point binomial. An expression for the ratio $\frac{\text{slope}}{\text{mean ordinate}}$ in terms of x is found, and the resulting differential equation solve for the relation of x and y .

The r th term of the symmetrical point binomial is

$$y_r = K \frac{n(n-1)\cdots(n-r+2)}{\underline{r-1}} \left(\frac{1}{2}\right)^n$$

The next term is

$$y_{r+1} = \frac{K n(n-1)\cdots(n-r+2)(n-r+1)}{\underline{r}} \left(\frac{1}{2}\right)^n$$

If these ordinates be plotted at equal intervals c along the x-axis, the slope of the line joining the tops of the ordinates

is
$$\frac{Y_{r+1} - Y_r}{c}$$

The mean ordinate is
$$\frac{Y_{r+1} + Y_r}{2}$$

$$\begin{aligned} \frac{\text{Slope}}{\text{Mean Ordinate}} &= \frac{Y_{r+1} - Y_r}{\frac{1}{2}(Y_{r+1} + Y_r)c} \\ &= \frac{n-r+1-r}{\frac{1}{2}(n-r+1+r)c} = \frac{c(n+2) - c(1+2r)}{\frac{1}{2}c^2(n+1)} \end{aligned}$$

But $cr = x_r$ and $c(r+1) = x_{r+1}$. And in a continuous curve the slope is $\frac{dy}{dx}$ and the mean ordinate is y .

$$\text{Then } \frac{1}{y} \frac{dy}{dx} = \frac{c(n+2) - (x_r + x_{r+1})}{\frac{1}{2}c^2(n+1)}$$

If x represent the mean abscissa as y the mean ordinate

$$\frac{1}{y} \frac{dy}{dx} = \frac{c(n+2) - 2x}{\frac{1}{2}c^2(n+1)} = \frac{\frac{c}{2}(n+2) - x}{\frac{1}{4}c^2(n+1)}$$

Making the linear substitution

$$x' = x - \frac{c}{2}(n+2)$$

which merely moves the y-axis $\frac{c}{2}(n+2)$ to the right and so,

of course, does not affect the shape of the curve, and writing

$$\sigma^2 = \frac{1}{4}c^2(n+1)$$

$$\frac{1}{y} \frac{dy}{dx} = -\frac{x}{\sigma^2}$$

Integrating

$$\begin{aligned} \log y &= -\frac{x^2}{2\sigma^2} + K \\ \therefore y &= C e^{-\frac{x^2}{2\sigma^2}} \end{aligned}$$

The generalized probability curve is derived from the asymmetrical point binomial in a strictly analogous manner. The r th term is

$$y_r = K \frac{n(n-1) \dots (n-r+2)}{r-1} p^{n-r+1} q^{r-1}$$

$$y_{r+1} = K \frac{n(n-1) \dots (n-r+2)(n-r+1)}{r} p^{n-r} q^r$$

$$\text{Slope} = \frac{y_{r+1} - y_r}{c}$$

$$\text{Mean ordinate} = \frac{1}{2} (y_{r+1} + y_r)$$

$$\text{Ratio} = \frac{y_{r+1} - y_r}{\frac{1}{2} c (y_{r+1} + y_r)} = \frac{\frac{n-r+1}{r} \cdot \frac{q}{p} - 1}{\frac{1}{2} c \left[\left(\frac{n-r+1}{r} \right) \frac{q}{p} + 1 \right]}$$

$$\text{Let } \frac{q}{p} = v$$

$$\text{Mean ordinate} = y$$

$$\text{Slope} = \frac{dy}{dx}$$

$$\begin{aligned} \text{Then } \frac{1}{y} \frac{dy}{dx} &= \frac{2}{c} \frac{(n-r+1)v - r}{(n-r+1)v + r} \\ &= \frac{2}{c} \frac{v(n+1) - r(1+L)}{v(n+1) - r(1-L)} \end{aligned} \quad (A)$$

$$\begin{aligned} \text{Let mean abscissa} &= x = \frac{1}{2} (x_r + x_{r+1}) \\ &= \frac{1}{2} [cr + c(r+1)] \\ 2x &= 2cr + c \\ r &= \frac{x}{c} - \frac{1}{2} \end{aligned}$$

Substituting this value of r in (A)

$$\frac{1}{y} \frac{dy}{dx} = \frac{2}{c} \frac{v(n+1) - (1+v)\left(\frac{x}{c} - \frac{1}{2}\right)}{v(n+1) + (1+v)\left(\frac{x}{c} - \frac{1}{2}\right)}$$

Replacing v by q/p and remembering $p+q=1$

$$\frac{1}{y} \frac{dy}{dx} = \frac{2}{c} \frac{cq(n+1) - (x - \frac{c}{2})}{cq(n+1) + (p-q)(x - \frac{c}{2})}$$

By moving y-axis by means of the linear substitution $x' = x - c(1/2 + q[n+1])$ and dropping primes

$$\frac{1}{y} \frac{dy}{dx} = \frac{2}{c} \frac{-x}{2cpq(n+1) + (p-q)x}$$

$$= \frac{2}{c(p-q)} \frac{-x}{\frac{2cpq(n+1)}{p-q} + x}$$

Let $K =$

$$\frac{2}{c(p-q)}$$

and $L =$

$$\frac{2cpq(n+1)}{p-q}$$

$$\frac{1}{y} \frac{dy}{dx} = K \left(\frac{-x}{x+L} \right) = -K + \frac{KL}{x+L}$$

Integrating

$$\log_e y = -Kx + KL \log_e (x+L) + \log_e y_0$$

B

$$\frac{y}{(x+L)^{KL}} = y_0 e^{-Kx}$$

$$y = y_0 (x+L)^{KL} e^{-Kx}$$

But for the problem solved $p = .601$, $q = .399$, $n = 2.632$,
 $c = 4.287$. Substituting these values in the expressions for
 K and L given above,

$$K = 2.310$$

$$L = 36.966$$

$$KL = 85.391$$

Hence everything is determined but y_0 which being a constant of integration must be determined by substituting simultaneous values of x and y . Such simultaneous values referred to the original axis ($1/2 c$ to the left of the initial point of the first rectangle of base c) and found on pp. 26 and 27.

$$\begin{array}{ll}
 x_1 = c = 4.287 & y_1 = 4993 \\
 x_2 = 2c = 8.574 & y_2 = 8723 \\
 x_3 = 3c = 12.851 & y_3 = 4726 \\
 x_4 = 4c = 17.148 & y_4 = 661
 \end{array}$$

But by change of axis, page ,

$$x' = x - c [1/2 + q(n+1)]$$

Substituting the values of c , q , and n shows the axis shifted to the right 8.3561 and hence falling at the age 10.0925.

The values of x then become

$$\begin{array}{ll}
 x_1 = -4.069 \\
 x_2 = .218 \\
 x_3 = 4.505 \\
 x_4 = 8.792
 \end{array}$$

As a check against errors, $\log_e y_0$ was computed three times by substituting the first three pairs of simultaneous values of x and y in $\log_e y$

$$\log_e y_0 = \log_e y + Kx - KL \log_e (x + L)$$

Whence

$$\begin{aligned}
 \log_{10} y_0 &= \log_{10} y + \frac{Kx}{2.3026} - KL \log_{10} (x + L) \\
 &= \log_{10} y + 1.0032 x - KL \log_{10} (x + L)
 \end{aligned}$$

The three values of $\log_{10} y_0$ so found are -129.936, -129.940, and -129.946. Using the average of these three gives

$$y = \log_{10}^{-1}(-129.941)(x + 36.966) 85.391 e^{-2.31 x}$$

Computing simultaneous values of x and y by this equation gives

x	Computed y	Average observed y	Difference	% error
-6	2431	985	1446	146.8
-5	3647	3433	214	6.2
-4	5003	6307	1504	20.7
-3	6403	7646	1243	16.3
-2	7570	7859	289	3.7
-1	8337	7924	413	5.2
0	8595	7857	538	6.8
1	8556	7666	690	9.0
2	7489	7362	127	1.7
3	6598	7030	432	6.1
4	5410	6164	754	12.2
5	4207	4633	426	9.2
6	3121	3050	71	2.3
7	2209	1858	351	18.9
8	1524	1080	444	40.8
9	<u>970</u>	581	389	<u>66.9</u>
Total	83865		Average =	23.5 %
Observed	<u>81743</u>			
Difference	2122 = 2.5 % of whole.		Average omitting end terms =	11.4 %

The x 's here tabulated are measured from an origin at the age 10.0925, and hence fall within accuracy of plotting at the integral ages 4, 5, - - - 19. The curve is continuous but the plotted data consisted of isolated points. If the number at any age as 4 be taken as one-half the sum of those included in the year-intervals immediately preceding and following there will be a possibility of comparing the ordinates of the probability curve with those of the observation polygon. The results are given in the table above. The fit

is not good and is the poorer as one recedes from the centroidal axis. The fact that the per cent error including the end terms including the end terms is 23.3 and excluding these terms is 11.4 emphasizes the source of error. This error stated in brief was in the failure of quadrature formulas. When the area of a figure is considered as concentrated at its centroid, the product of the area and the distance of this centroid from the axis of moments gives the first moment exactly, for the centroid is defined as the point for which

$$\bar{x} = \frac{\int xy \, dx}{\int y \, dx}$$

But when the second moment is computed with the square of this same distance as a multiplier for the area this assumes

$$\bar{x} = \text{radius of gyration}$$

or

$$\frac{\int xy \, dx}{\int y \, dx} = \left[\frac{\int x^2 y \, dx}{\int y \, dx} \right]^{1/2}$$

which is not true. And in the higher moments this same sort of error enters to a still larger extent, for an error in the x affects the result the more seriously as the x is raised to a higher power. That this is true and that the error will be greater in terms farthest from the axis of moments may be shown by

$$(x + x')^4 - x^4 = 4x^3x' + 6x^2(x')^2 + 4x(x')^3 + (x')^4$$

where x is the true value and x' the error.

This problem will perhaps be useful as an illustration of one of the type curves used by Prof. Pearson, full treatment of which may be found in Philosophical Transactions of the Royal Society, volume 186A, part 1, pages 343 - 414. Prof. Pearson seems to have fairly well settled the question as to the relative merits of the method of moments and that of least squares for statistical problems, and has shown conclusively that his various types of curves derived from the point binomial may be fitted by the method of moments to large classes of observation curves where the method of least squares has nothing better or more flexible to offer than the normal frequency curve.

QUADRATURE FORMULAS.

Since the failure to obtain an accurately fitting curve in the problem solved is manifestly due to lack of quadrature formulas, a section of this paper will be devoted to discussing some common formulas and their suitability to the matter in hand. The treatment is in terms of the calculus of finite differences, and the notation used is as follows:

$w_x, w_{x+1}, \dots$ = any function of x , the same function of $x+1, \dots$ of $x+1, \dots$

Dw_x = the increment in the function due to a unit increase in $x = w_{x+1} - w_x$.

D^2w_x = result of repeating operation denoted by D

$$\begin{aligned} &= D(Dw_x) = Dw_{x+1} - Dw_x \\ &= w_{x+2} - w_{x+1} - [w_{x+1} - w_x]. \end{aligned}$$

To illustrate, let $w_x = x^3$.

Taking $x = 1, 2, \dots, 8$

$w_x, w_{x+1}, \dots, w_{x+7}$	=	1	8	27	64	125	216	343	512
$Dw_x, Dw_{x+1}, \dots, Dw_{x+6}$	=	7	19	37	61	91	127	169	
$D^2w_x, D^2w_{x+1}, \dots, D^2w_{x+5}$	=	12	18	24	30	36	42		
$D^3w_x, D^3w_{x+1}, \dots, D^3w_{x+4}$	=	6	6	6	6	6			
$D^4w_x, D^4w_{x+1}, \dots, D^4w_{x+3}$	=	0	0	0	0				

The abbreviated notation D_r^n is used sometimes for D_{x+r}^n

By definition,

$$\begin{aligned}
 D_r^n &= D_{r-1}^n + D_{r-1}^{n+1} \\
 &= (D_{r-2}^n + D_{r-2}^{n+1}) + (D_{r-2}^{n+1} + D_{r-2}^{n+2}) \\
 &= D_{r-2}^n + 2D_{r-2}^{n+1} + D_{r-2}^{n+2} \\
 &= D_{r-3}^n + 3D_{r-3}^{n+1} + 3D_{r-3}^{n+2} + D_{r-3}^{n+3} \\
 &\quad - - - - - \\
 &= D_0^n + nD_0^{n+1} + \frac{n(n-1)}{2!} D_0^{n+2} + - - - + D_0^{n+r}
 \end{aligned}$$

as may be easily shown by induction or seen by inspection since the coefficients are formed by the same succession of additions as those in finding the n th power of a binomial.

Also by definition

$$w_{x+1} = w_x + D_0$$

$$w_{x+2} = w_{x+1} + D_1$$

$$= w_x + 2D_0 + D_0^2$$

$$w_{x+3} = w_{x+2} + D_2$$

$$= w_x + 3D_0 + 3D_0^2 + D_0^3$$

Hence from the way in which the coefficients result from successive additions

$$\begin{aligned}w_{x+r} &= w_x + rD_0 + \frac{r(r-1)}{2!} D_0^2 + - - - + D_0^r \\&= w_x + rDw_x + \frac{r(r-1)}{2!} D^2w_x + - - - + D^r w_x\end{aligned}$$

If $x = 0$ and $r = x$

$$w_x = w_0 + xDw_0 + \frac{x(x-1)}{2!} D^2w_0 + \frac{x(x-1)(x-2)}{3!} D^3w_0$$

a form which has been called the conjugate of Maclaurin's theorem.

Suppose then the problem given is to compute the area under a curve from $n+1$ equidistant ordinates.

Let it be assumed as in the original development of the method of moments that there exists an equation

$$w_x = 0$$

which exactly fits the curve in question and whose expansion by Taylor's theorem converges fairly rapidly. As will be shown in the section on the choice of form of function to fit given data, this is equivalent to saying that the successive orders of differences diminish rapidly.

Let the y-axis coincide with the first ordinate. Then the successive ordinates are $w_0, w_1, w_2, - - - w_n$.

By the first part of this paragraph,

$$w_x = w_0 + xDw_0 + \frac{x(x-1)D^2w_0}{2!} + \frac{x(x-1)(x-2)D^3w_0}{3!} + \dots$$

and the area under the curve is

$$\begin{aligned} \int_0^n w_x dx &= w_0 \int_0^n dx + Dw_0 \int_0^n x dx + \frac{D^2w_0}{2!} \int_0^n (x^2 - x) dx \\ &\quad + \frac{D^3w_0}{3!} \int_0^n (x^3 - 3x^2 + 2x) dx + \frac{D^4w_0}{4!} \int_0^n (x^4 - 6x^3 + 11x^2 - 6x) dx \\ &\quad + \frac{D^5w_0}{5!} \int_0^n (x^5 - 10x^4 + 35x^3 - 50x^2 + 24x) dx \\ &\quad + \frac{D^6w_0}{6!} \int_0^n (x^6 - 15x^5 + 35x^4 - 225x^3 + 274x^2 - 120x) dx + \dots \\ &= nw_0 + \frac{n^2}{2} Dw_0 + \left(\frac{n^3}{3} - \frac{n^2}{2}\right) \frac{D^2w_0}{2!} + \left(\frac{n^4}{4} - n^3 + n^2\right) \frac{D^3w_0}{3!} \\ &\quad + \left(\frac{n^5}{5} - \frac{3n^4}{2} + \frac{11n^3}{3} - 3n^2\right) \frac{D^4w_0}{4!} + \left(\frac{n^6}{6} - 2n^5 + \frac{35n^4}{4} - \frac{50n^3}{3} \right. \\ &\quad \left. + 12n^2\right) \frac{D^5w_0}{5!} \\ &\quad + \left(\frac{n^7}{7} - \frac{5n^6}{2} + 17n^5 - \frac{225n^4}{4} + \frac{274n^3}{3} - 60n^2\right) \frac{D^6w_0}{6!} + \dots \end{aligned}$$

If only 2 ordinates are given the area is a trapezoid and the area would ordinarily be computed by $\frac{w_0 + w_1}{2}$. Applying this general formula gives since $n = 1$ and $D^2w_0 = 0$

$$\begin{aligned} \int_0^1 w_x dx &= w_0 + 1/2 Dw_0 \\ &= w_0 + 1/2(w_1 - w_0) \\ &= \frac{w_0 + w_1}{2} \end{aligned}$$

For a series of adjacent trapezoids each ordinate except the first and last must be counted in two trapezoids and hence

$$\text{Area} = 1/2 (w_0 + w_l) + w_1 + w_2 + w_3 + \dots + w_{l-1}$$

where $l+1$ is the number of ordinates. If the ordinates are at distances h instead of unity

$$I \int z \, dx = \frac{h}{2} [z_0 + z_l + 2(z_1 + z_2 + z_3 + \dots + z_{l-1})]$$

where x represents successively $y, xy, x^2y, \dots, x^n y$ and z_r is the value of z when $x = rh$.

If only three ordinates are known $n = 2$ and $D^3 w_0 = 0$ and hence the area is

$$\begin{aligned} \int_0^2 w_x dx &= 2w_0 + 2Dw_0 + \frac{(8-2)}{3} \frac{D^2 w_0}{2!} \\ &= 2w_0 + 2Dw_0 + \frac{1}{3} D^2 w_0 \\ &= 2w_0 + 2(w_1 - w_0) + \frac{1}{3} (w_2 - 2w_1 + w_0) \\ &= \frac{w_0 + 4w_1 + w_2}{3} \end{aligned}$$

This is of course equivalent to assuming a second degree parabola passed through the three points and finding the area under this parabola. It is the common form of Simpson's rule, and is readily extended to any figure where an odd number of ordinates

are given, for there are then an even number of elementary areas and if they be grouped in 2's the total area will count the first and last ordinates once each, the second, fourth, sixth, etc. four times each as a mid-ordinate, and the third, fifth, seventh, etc. each twice, once as a last ordinate in the preceding area, once as initial ordinate of the succeeding area. If the ordinates are spaced at a distance h from each other instead of at unit distance, Simpson's rule becomes, for $2p+1$ ordinates,

$$\text{II} \quad \int z \, dx = \frac{h}{3} \left[z_0 + z_{2p} + 4(z_1 + z_3 + z_5 + \dots + z_{2p-1}) + 2(z_2 + z_4 + \dots + z_{2p-2}) \right]$$

It is, in general, a better form for quadrature than the assumption that the area is a series of trapezoids just as a smooth curve through a set of points is a better representation of their law than the straight line polygon which they determine.

If four ordinates are given $n = 3$ and $D^4 w_0 = 0$.

The area then becomes

$$\begin{aligned} \int_0^3 w \, dx &= 3w_0 + \frac{9}{2} D w_0 + \frac{(9-9)}{2} \frac{D^2 w_0}{2!} + \frac{(81-27+9)}{4} \frac{D^3 w_0}{3} \\ &= 3w_0 + \frac{9}{2}(w_1 - w_0) + \frac{9}{4}(w_2 - 2w_1 + w_0) + \frac{3}{8}(w_3 - 3w_2 + 3w_1 - w_0) \\ &= \frac{3}{8}w_0 + \frac{9}{8}w_1 + \frac{9}{8}w_2 + \frac{3}{8}w_3 = \frac{3}{8}(w_0 + 3w_1 + 3w_2 + w_3) \end{aligned}$$

This result is Newton's rule, is perfectly accurate for third

degree parabolas and is in general for any ordinary function a closer approximation than Simpson's rule. It has the disadvantage of applying only to those cases where the number of elementary areas is a multiple of 3. With the same notation as used in generalizing Simpson's Rule, this gives for a number of elementary areas bounded by $3p+1$ ordinates and placed at equal distances h ,

$$\text{III} \quad \int z \, dx = \frac{3h}{8} \left[z_0 + z_{3p} + 3(z_1 + z_2 + z_4 + z_5 + z_7 + z_8 - - z_{3p-1}) \right. \\ \left. 2(z_3 + z_6 + - - - + z_{3p-3}) \right]$$

If five ordinates are given, $n = 4$ and $D^5 w_0 = 0$.

Then the area

$$\begin{aligned} \int_0^4 w_x \, dx &= 4w_0 + 8Dw_0 + \left(\frac{64-8}{3}\right) \frac{D^2 w_0}{2!} + (64-64+16) \frac{D^3 w_0}{3!} \\ &\quad + \left(\frac{1024-384+704-48}{5}\right) \frac{D^4 w_0}{4!} \\ &= 4w_0 + 8(w_1 - w_0) + \frac{20}{3}(w_2 - 2w_1 + w_0) + \frac{8}{3}(w_3 - 3w_2 + 3w_1 - w_0) \\ &\quad + \frac{14}{45}(w_4 - 4w_3 + 6w_2 - 4w_1 + w_0) \\ &= \frac{14(w_0 + w_4) + 64(w_1 + w_3) + 24w_2}{45} \end{aligned}$$

Generalizing as before for $4p+1$ ordinates at equal distances h .

$$\text{IV } \int z \, dx = \frac{2h}{45} \left[7(z_0 + z_{4p}) + 32(z_1 + z_3 + z_5 + z_7 + \dots + z_{4p-1}) \right. \\ \left. + 12(z_2 + z_6 + z_{10} + \dots + z_{4p-2}) \right. \\ \left. + 14(z_4 + z_8 + z_{12} + \dots + z_{4p-4}) \right]$$

The form taken by this formula when $n = 6$ is especially good for data like those fitted by a sine curve, where no order of differences vanishes but they do diminish in size rather rapidly.

$$\int_0^6 w_x \, dx = 6w_0 + 18Dw_0 + (72-18)\frac{D^2w_0}{2!} + (324-216+36)\frac{D^3w_0}{3!} \\ + \frac{(7776-1944+792-108)}{5}\frac{D^4w_0}{4!} \\ + (7776-15552+11340-3600+432)\frac{D^5w_0}{5!} \\ + \frac{(279936-116640+132192-72900+19728-2160)}{7}\frac{D^6w_0}{6!} \\ = 6w_0 + 18Dw_0 + 27D^2w_0 + 24D^3w_0 + \frac{123}{10}D^4w_0 + \frac{33}{10}D^5w_0 + \frac{41}{140}D^6w_0$$

Since D^6w_0 is small a very slight error is introduced, while the labor in application is much reduced, by substituting for the last coefficient $\frac{42}{140} = \frac{3}{10}$. The formula then becomes

$$\text{Area} = \frac{3}{10} [w_0 + w_2 + w_4 + w_6 + 5(w_1 + w_5) + 6w_3]$$

It is a peculiar fact that the coefficient of D^7w_0 vanishes i-



dentically, for the quantity to be integrated (page 38) is

$$\begin{aligned} & \int_0^6 x(x-1) - - - (x-6) \frac{D^7 w_0}{7!} dx \\ & \frac{D^7 w_0}{7!} \int_0^6 (x^7 - 21x^6 + 175x^5 - 735x^4 + 1624x^3 - 1764x^2 + 720x) \\ & = \frac{D^7 w_0}{7!} \left[\frac{n^8}{8} - 3n^7 + \frac{175n^6}{6} - 147n^5 + 406n^4 - 588n^3 + 360n^2 \right]_{n=6} \end{aligned}$$

Putting 6 for n the term reduces to zero. Then

$$\frac{3}{10} [w_0 + w_2 + w_4 + w_6 + 5(w_1 + w_5) + 6w_3]$$

gives the area exactly if the fifth order of differences is constant and with an error of only $\frac{1}{140} D^6 w_0$ when either sixth or seventh order of differences is constant.

Generalizing for $6p+1$ ordinates at equal distances h

$$\begin{aligned} V \int z \, dx &= \frac{3h}{10} [z_0 + z_2 + z_4 + z_6 + z_8 + z_{10} + - - - + z_{6p} \\ &+ 5(z_1 + z_5 + z_7 + z_{11} + - - - + z_{6p-1}) \\ &+ 6(z_3 + z_9 + z_{15} + - - - + z_{6p-3}) + 2(z_6 + z_{12} + - - - + z_{6p-6})] \end{aligned}$$

This formula is called Weddle's Formula and is of much use and great accuracy in quadratures where there is really a mathematical law expressed in the data as in case there actually is a sine curve whose constants are to be determined by a few isolated points. But manifestly in fitting a set of very irregularly placed points which represent for instance a statistical

problem, where there is no reason to expect continuity and where the irregularity may be due to the limited number of observations, this formula, which in effect assumes a sixth degree parabola through each group of seven points is no more likely to be true than the assumption of a succession of trapezoids. All the formulas of this article are subject to the disadvantage that they give to certain ordinates and hence to certain observations greater weight in the determination of the final result than to others. Thus all of them make the end points less important in determining the curve, and the result is that the end points are not so closely fitted as the others.

THE PARABOLIC CURVE.

$$y = a_0 + a_1x + a_2x^2 + a_3x^3 + \dots + a_nx^n$$

has been so universally accepted as a suitable function to represent most engineering and many other formulas because of its simplicity and because of the readiness with which the method of least squares can be applied to it. It is not the purpose of this paper to advocate the indiscriminate use of this function, which the writer thinks has been much used where other functions would answer the purpose better, but it can be shown that even here where the method of least squares is best adapted to the end sought, the method of moments is better. It will be remembered that theoretically one is as good as the other because the method of moments for functions of this form is not an approximation but is absolutely accurate. The advantage lies in the fact that much of the work can be done once for all and the results applied with a reduction of the labor involved to the solution of the individual problem in hand.

Let the assumed function be

$$y = y_0 \left[e_0 + e_1 \left(\frac{x}{l} \right) + e_2 \left(\frac{x}{l} \right)^2 + \dots + e_{n-1} \left(\frac{x}{l} \right)^{n-1} \right]$$

where $2l$ is the total range of observations along the axis of x .

$$A = \text{area} = 2l y_0$$

$$\therefore y_0 = \frac{A}{2l}$$

Assume the origin at the middle of the range, then for even moments

$$\begin{aligned} \frac{AM_{2r}}{l^{2r}} &= \int_{-l}^l \frac{yx^{2r}}{l^{2r}} dx = y_0 \left[\frac{e_0}{2r+1} \left(\frac{x^{2r+1}}{l^{2r}} \right) + \frac{e_1}{2r+2} \left(\frac{x^{2r+2}}{l^{2r+1}} \right) + \dots + \frac{e_{n-1}}{2r+1} \left(\frac{x^{2r+n}}{l^{2r+n-1}} \right) \right]_{-l}^l \\ &= 2ly_0 \left[\frac{e_0}{2r+1} \left(\frac{x}{l} \right)^{2r+1} + \frac{e_1}{2r+2} \left(\frac{x}{l} \right)^{2r+2} + \dots + \frac{e_{n-1}}{2r+1} \left(\frac{x}{l} \right)^{2r+n} \right]_{-l}^l \\ &= 2ly_0 \left[\frac{e_0}{2r+1} + \frac{e_2}{2r+3} + \frac{e_4}{2r+5} + \dots \right] \end{aligned}$$

for substituting the limits l and $-l$ make even powers of x cancel and give twice those terms containing odd powers.

$$\therefore \frac{AM_{2r}}{l^{2r}} = A \left[\frac{e_0}{2r+1} + \frac{e_2}{2r+3} + \frac{e_4}{2r+5} + \dots \right]$$

$$\text{I} \quad \therefore \frac{M_{2r}}{l^{2r}} = \frac{e_0}{2r+1} + \frac{e_2}{2r+3} + \frac{e_4}{2r+5} + \dots$$

For odd moments similar reasoning gives

$$\begin{aligned} \frac{AM_{2r+1}}{l^{2r+1}} &= \int_{-l}^l \frac{yx^{2r+1}}{l^{2r+1}} dx = ly_0 \left[\frac{e_0}{2r+2} \left(\frac{x}{l} \right)^{2r+2} + \frac{e_1}{2r+3} \left(\frac{x}{l} \right)^{2r+3} + \frac{e_2}{2r+4} \left(\frac{x}{l} \right)^{2r+4} + \dots \right] \\ &= 2ly_0 \left[\frac{e_1}{2r+3} + \frac{e_3}{2r+5} + \frac{e_5}{2r+7} + \dots \right] \end{aligned}$$

$$\text{II} \quad \therefore \frac{M_{2r+1}}{l^{2r+1}} = \frac{e_1}{2r+3} + \frac{e_3}{2r+5} + \frac{e_5}{2r+7} + \dots$$

Let $L_s = \frac{M_s}{z^s}$. Then from I

$$L_0 = \frac{M_0}{z^0} = \frac{1}{1} = 1 = e_0 + \frac{e_2}{3} + \frac{e_4}{5} + \frac{e_6}{7} + \dots$$

$$L_2 = \frac{M_2}{z^2} = \frac{e_0}{3} + \frac{e_2}{5} + \frac{e_4}{7} + \frac{e_6}{9} + \dots$$

$$L_4 = \frac{M_4}{z^4} = \frac{e_0}{5} + \frac{e_2}{7} + \frac{e_4}{9} + \frac{e_6}{11} + \dots$$

From II

$$L_1 = \frac{M_1}{z} = \frac{e_1}{3} + \frac{e_3}{5} + \frac{e_5}{7} + \frac{e_7}{9} + \dots$$

$$L_3 = \frac{M_3}{z^3} = \frac{e_1}{5} + \frac{e_3}{7} + \frac{e_5}{9} + \frac{e_7}{11} + \dots$$

Then for a second order parabola,

$$L_0 = e_0 + \frac{e_2}{3}$$

$$L_1 = \frac{e_1}{3}$$

$$L_2 = \frac{e_0}{3} + \frac{e_2}{5}$$

Solving for the e 's

III

$$e_0 = \frac{3}{4}(3L_0 - 5L_2)$$

$$e_1 = 3L_1$$

$$e_2 = \frac{15}{4}(3L_2 - L_0)$$

By the use of III the constants may be computed directly as soon as the area and first two moments of the observation curve have been determined.

For the third degree parabola,

$$y = y_0 \left[e_0 + e_1 \frac{x}{l} + e_2 \left(\frac{x}{l} \right)^2 + e_3 \left(\frac{x}{l} \right)^3 \right]$$

$$L_0 = e_0 + \frac{e_2}{3}$$

$$L_1 = \frac{e_1}{3} + \frac{e_3}{5}$$

$$L_2 = \frac{e_0}{3} + \frac{e_2}{5}$$

$$L_3 = \frac{e_1}{5} + \frac{e_3}{7}$$

Whence

$$e_0 = \frac{3}{4}(3L_0 - 5L_2)$$

$$e_1 = \frac{15}{4}(5L_1 - 7L_3)$$

IV

$$e_2 = \frac{15}{4}(3L_2 - L_0)$$

$$e_3 = \frac{35}{4}(5L_3 - 3L_1)$$

For the fourth degree parabola

$$y = y_0 \left[e_0 + e_1 \frac{x}{l} + e_2 \left(\frac{x}{l} \right)^2 + e_3 \left(\frac{x}{l} \right)^3 + e_4 \left(\frac{x}{l} \right)^4 \right]$$

$$L_0 = e_0 + \frac{e_2}{3} + \frac{e_4}{5}$$

$$L_1 = \frac{e_1}{3} + \frac{e_3}{5}$$

$$L_2 = \frac{e_0}{3} + \frac{e_2}{5} + \frac{e_4}{7}$$

$$L_3 = \frac{e_1}{5} + \frac{e_3}{7}$$

$$L_4 = \frac{e_0}{5} + \frac{e_2}{7} + \frac{e_4}{9}$$

Whence

$$e_0 = \frac{15}{64}(15L_0 - 70L_2 + 63L_4)$$

$$e_1 = \frac{15}{4}(5L_1 - 7L_3)$$

$$V \quad e_2 = \frac{105}{32}(-5L_0 + 42L_2 - 45L_4)$$

$$e_3 = \frac{35}{4}(5L_3 - 3L_1)$$

$$e_4 = \frac{315}{64}(3L_0 - 50L_2 + 35L_4)$$

For the fifth degree parabola,

$$y = y_0 \left(e_0 + e_1 \frac{x}{2} + e_2 \left(\frac{x}{2} \right)^2 + e_3 \left(\frac{x}{2} \right)^3 + e_4 \left(\frac{x}{2} \right)^4 + e_5 \left(\frac{x}{2} \right)^5 \right)$$

$$e_0 = \frac{15}{64}(15L_0 - 70L_2 + 63L_4)$$

$$e_1 = \frac{105}{32}(33L_1 - 126L_3 + 99L_5)$$

$$e_2 = \frac{105}{32}(-5L_0 + 42L_2 - 45L_4)$$

$$e_3 = \frac{315}{64}(-21L_1 + 90L_3 - 77L_5)$$

VI

$$e_4 = \frac{315}{64}(3L_0 - 50L_2 + 35L_4)$$

$$e_5 = \frac{693}{64}(15L_1 - 70L_3 + 63L_5)$$

For the sixth degree parabola

$$e_0 = \frac{35}{256}(35L_0 - 315L_2 + 393L_4 - 429L_6)$$

$$e_1 = \frac{105}{64}(35L_1 - 126L_3 + 99L_5)$$

$$e_2 = \frac{315}{256}(-35L_0 + 567L_2 - 1485L_4 + 1001L_6)$$

$$\text{VII } e_3 = \frac{315}{32}(-21L_1 + 90L_3 - 77L_5)$$

$$e_4 = \frac{3435}{256}(7L_0 - 135L_2 + 385L_4 - 273L_6)$$

$$e_5 = \frac{693}{64}(15L_1 - 70L_3 + 63L_5)$$

$$e_6 = \frac{3003}{256}(-5L_0 + 105L_2 - 315L_4 + 231L_6)$$

Illustrative problem on velocities of the Mississippi River at different depths. Let it be required to fit a second degree parabola to the following data.

x	y	x	y	x	y
0	3.1950	.3	3.2611	.7	3.1266
.1	3.2299	.4	3.2516	.8	3.0594
.2	3.2532	.5	3.2282	.9	2.9759
		.6	3.1807		

where x = the fraction of the depth at which the observations

were taken. y = observed velocity in feet per second.

Using Newton's Rule for quadrature

$$\int z \, dx = \frac{3h}{8} [z_0 + z_{5p} + 3(z_1 + z_2 + z_4 + z_5 + \dots) + 2(z_3 + z_6 + \dots)]$$

the work is tabulated as follows. Columns marked x and y contain the data, those marked xy and x^2y products of corresponding values of the variables. The column $S(y)$ contains the successive terms of the y column multiplied by the numerical factor which is assigned to it by Newton's formula above, that is, the first by 1, next two by 3, next by 2, etc. Hence the sum of this column multiplied by $3/8 h$ gives the total area. The column $S(xy)$ and $S(x^2y)$ in the same way give the first and second moments.

x	y	$S(y)$	xy	$S(xy)$	x^2y	$S(x^2y)$
0	3.1950	3.1950	0	0	0	0
.1	3.2299	9.6897	.32299	.96897	.032299	.096897
.2	3.2532	9.7596	.65064	1.95192	.130128	.390384
.3	3.2611	6.5222	.97833	1.95666	.293499	.586998
.4	3.2516	9.7548	1.30064	3.90192	.520256	1.560768
.5	3.2282	9.6846	1.61410	4.84230	.807050	2.421150
.6	3.1807	6.5614	1.90842	3.81684	1.145052	2.290104
.7	3.1266	9.3798	2.18862	6.56586	1.532034	4.596102
.8	3.0594	9.1782	2.44752	7.34256	1.958016	5.874048
.9	2.9759	<u>2.9759</u>	2.67831	<u>2.67831</u>	2.410479	<u>2.410479</u>
		76.5012		34.02534		20.226930

$$AM_0 = \frac{3h}{8} \times 76.5012 \quad h = \frac{1}{10}$$

$$= \frac{3}{80} \times 76.5012 = 2.868795 = A$$

$$AM_1 = \frac{3}{80} \times 34.0254 = 1.275950$$

$$M_1 = .444769$$

$$AM_2 = \frac{3}{80} \times 20.226930 = .7585099$$

$$M_2 = .264400$$

But these M's are calculated about a y-axis at one end of the range, and in order to apply formulas V, the M's about an axis through the mid-point of the range are required. These might of course have been calculated originally but since the mid-point comes at $x = .45$ each x would have been expressed in two figures and the arithmetical work thereby increased. The axis of moments must now be shifted to this mid-point, a distance of .45. Hence by the formula

$$M'_n = M_n - dnM_{n-1} + d^2 \frac{n(n-1)}{2!} M_{n-2} + \dots$$

derived on page 20 ,

$$M'_0 = M_0 = 1$$

$$M'_1 = M_1 - .45 = -.005231$$

$$M'_2 = M_2 - .9M_1 + .2025 = .066608$$

$$2 \text{ l} = \text{the range} = .9$$

$$l = .45$$

Curve showing relation of
Velocity and Depth

in

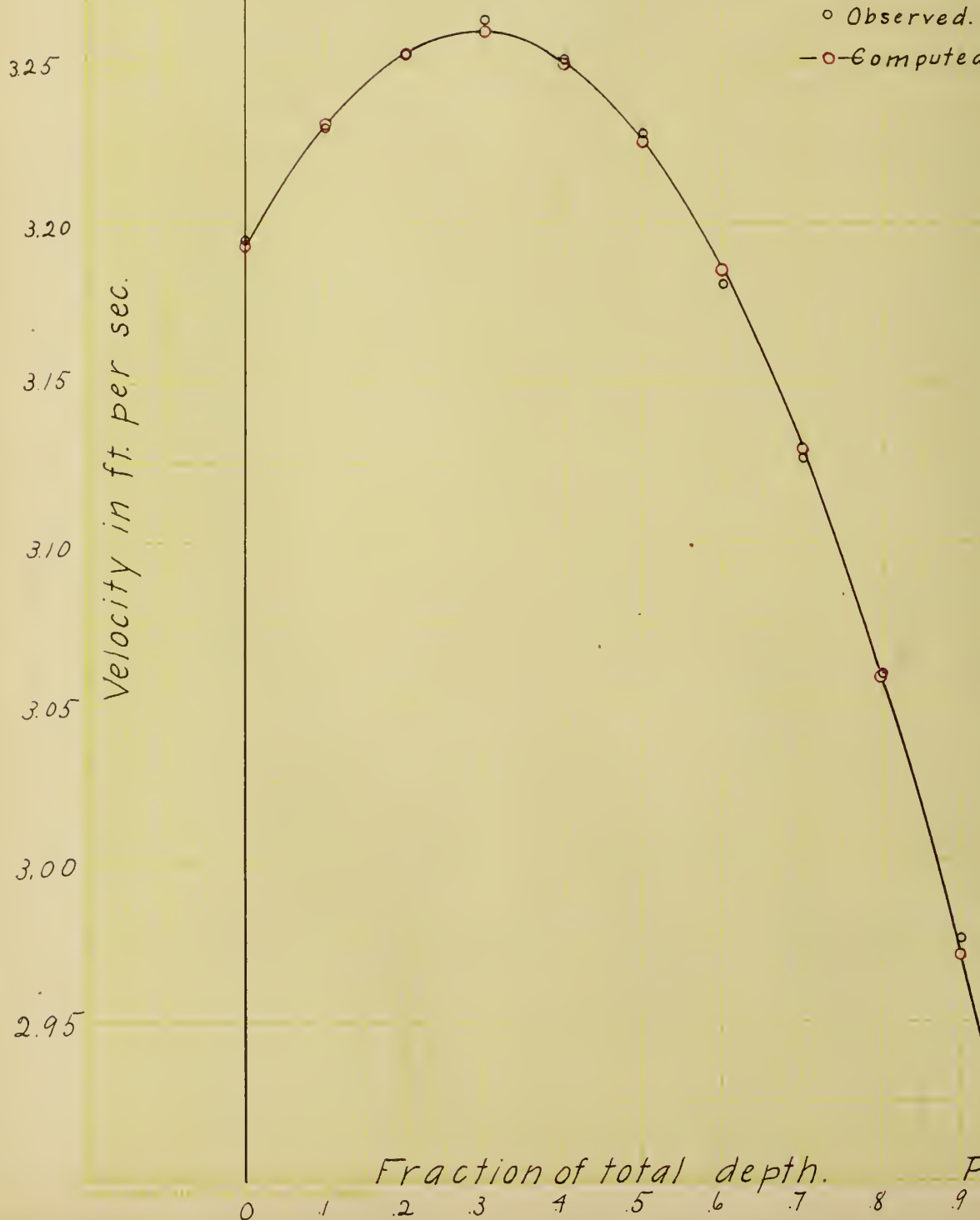
Mississippi River.

$$Y = 3.193389 + .455099X - .780123X^2$$

Method of Moments.

○ Observed.

—○— Computed.



$$y_0 = \frac{A}{2 \cdot 1} = 3.187550$$

$$L_0 = \frac{M_0}{1} = 1$$

$$L_1 = \frac{M_1}{1} = .011624$$

$$L_2 = \frac{M_2}{1^2} = .328928$$

$$e_0 = \frac{3}{4}(3L_0 - 5L_2) = 1.016520$$

$$e_1 = 3L_1 = -.034872 \quad \text{By formula III.}$$

$$e_2 = \frac{15}{4}(3L_2 - L_0) = -.049560$$

Hence the equation of the curve referred to an axis through the middle of the range is

$$y = 3.187550(1.016520 - \frac{.034872}{.45} x - \frac{.049560}{.2025} x^2)$$

Transferring back to original axis in order to have a more convenient form of equation by using the substitution

$$x = x' - .45$$

$$y = 3.18755(1.001832 + .142774 x - .244741 x^2)$$

$$= 3.193389 + .455099 x - .780123 x^2$$

The same data will now be fitted by a function of the same form by the method of least squares and the amount of work involved in the two methods compared.

$$\text{Assume } y = a + bx + cx^2$$

Substituting simultaneous values of x and y

$$3.1950 = a + 0. b + 0. c$$

$$3.2299 = a + .1b + .01c$$

$$3.2532 = a + .2b + .04c$$

$$3.2611 = a + .3b + .09c$$

$$3.2516 = a + .4b + .16c$$

Observation equations.

$$3.2282 = a + .5b + .25c$$

$$3.1807 = a + .6b + .36c$$

$$3.1266 = a + .7b + .49c$$

$$3.0594 = a + .8b + .64c$$

$$\underline{2.9759} = \underline{a} + \underline{.9b} + \underline{.81c}$$

$$I \quad 31.7616 = 10a + 4.5b + 2.85c$$

I is the first normal equation for all the coefficients of a are unity.

Multiplying each equation through by the coefficient of b and forming second normal equation gives

$$.32299 = .1a + .01b + .001c$$

$$.65064 = .2a + .04b + .008c$$

$$.97833 = .3a + .09b + .027c$$

$$1.30064 = .4a + .16b + .064c$$

$$1.61410 = .5a + .25b + .125c$$

$$1.90842 = .6a + .36b + .216c$$

$$2.18862 = .7a + .49b + .343c$$

$$2.44752 = .8a + .64b + .512c$$

$$\underline{2.67831} = \underline{.9a} + \underline{.81b} + \underline{.729c}$$

$$\text{II} \quad 14.08957 = 4.5a + 2.85b + 2.025c$$

Multiplying through by coefficients of c and forming the third normal equation gives

$$.032299 = .01a + .001b + .0001c$$

$$.130128 = .04a + .008b + .0026c$$

$$.293499 = .09a + .027b + .0081c$$

$$.520256 = .16a + .064b + .0256c$$

$$.807050 = .25a + .125b + .3125c$$

$$1.145052 = .36a + .216b + .1296c$$

$$1.532034 = .49a + .343b + .2301c$$

$$1.058016 = .64a + .512b + .4096c$$

$$\underline{2.410479} = \underline{.81a} + \underline{.729b} + \underline{.6561c}$$

$$\text{III} \quad 8.828813 = 2.85a + 2.925b + 1.5333c$$

The three equations

$$\text{I} \quad 31.7616 = 10a + 4.5b + 2.85c$$

$$\text{II} \quad 14.08957 = 4.5a + 2.85b + 2.025c$$

$$\text{III} \quad 8.828813 = 2.85a + 2.925b + 1.5333c$$

must be solved for a, b and c.

$$\text{I} \quad 28.58544 = 9a + 4.05b + 2.565c$$

$$\text{II} \quad \underline{28.17914} = \underline{9a} + \underline{5.70b} + \underline{4.050c}$$

$$\text{IV} \quad -.4063 = 1.65b + 1.485c$$

Curve showing relation of
Velocity and Depth
in

Mississippi River.

$$V = 3.19513 + 4.4253x - .7653x^2$$

Least Squares.

○ Observed.

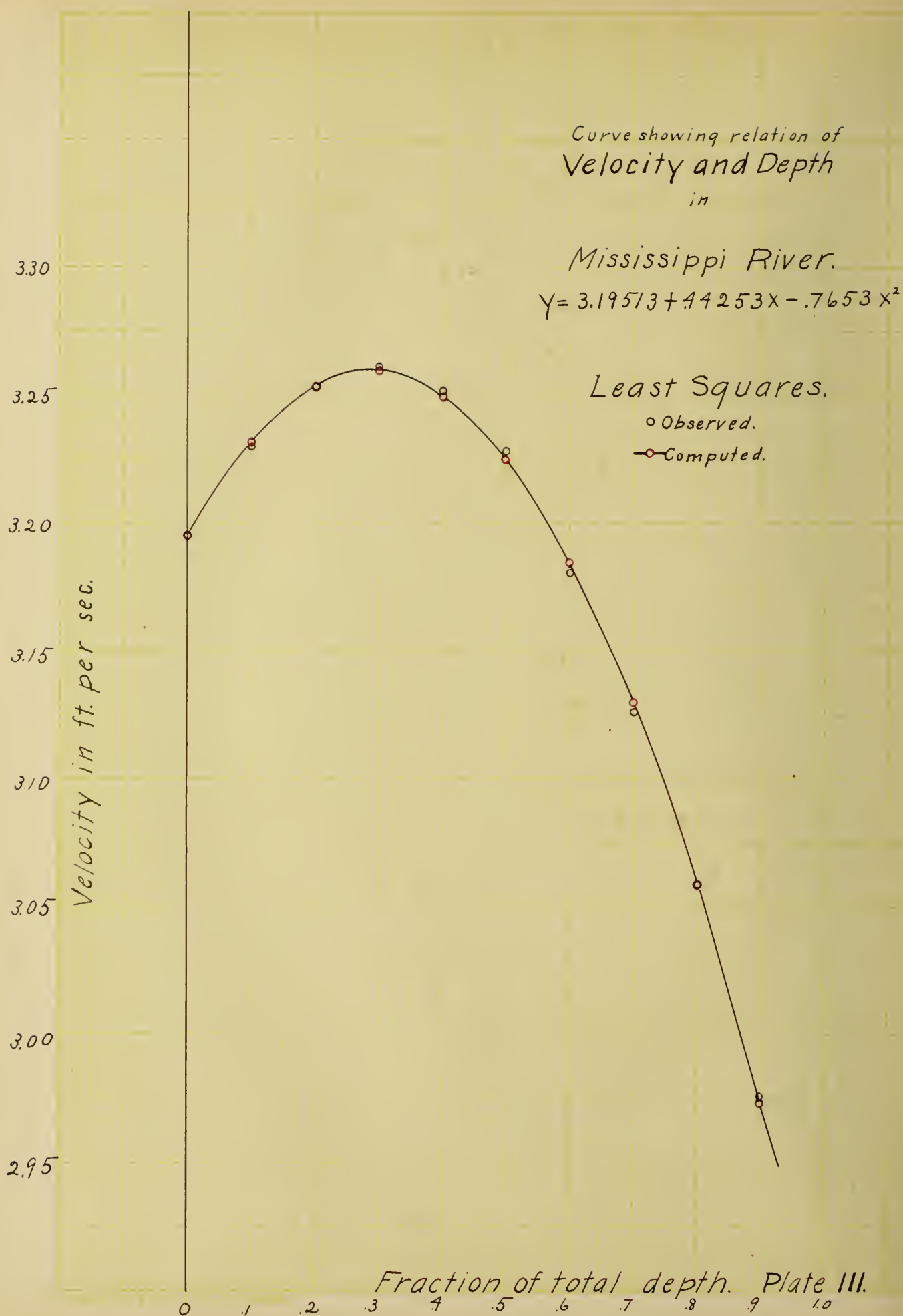
—○— Computed.

3.30
3.25
3.20
3.15
3.10
3.05
3.00
2.95

Velocity in ft. per sec.

Fraction of total depth. Plate III.

0 .1 .2 .3 .4 .5 .6 .7 .8 .9 1.0



$$\text{II} \quad 26.486439 = 8.55a + 6.075b + 4.5999c$$

$$\text{III} \quad \underline{26.770183 = 8.55a + 5.415b + 3.8475c}$$

$$\text{V} \quad -.283744 = .660b + .7524c$$

$$\text{IV} \quad - .8126 = 3.3b + 2.97c$$

$$\text{V} \quad \underline{-1.41872 = 3.3b + 3.762c}$$

$$\text{VI} \quad - .60612 = .792c$$

$$c = - .7653$$

$$\text{IV} \quad - .4063 = 1.65b - 1.1364715$$

$$b = .44253$$

$$\text{I} \quad 31.7616 = 10a + 1.991385 - 2.181105$$

$$a = 3.19513$$

Hence the equation is

$$y = 3.19513 + .44253 x - .7653 x^2$$

All the work for solving this problem by both methods has been given in order that the amount of labor involved might be compared in a problem where all the possible advantages were on the side of the least squares method. Here it was required to transfer the axes in the final result to a certain position, namely, with the origin on the surface of the water. This was not a necessity inherent in the problem of curve-fitting but was necessary for convenience in using the formula

and for comparison of results of the two methods. It will be seen that even with this additional labor the method of moments is much the less laborious method. It remains to tabulate the relative accuracy of the results.

Observed		Least squares			Moments		
x	y	y	v	v ²	y	v	v ²
0	3.1959	3.1951	-.0001	.000000	3.1934	.0016	.000003
.1	3.2299	3.2317	-.0018	3	3.2311	-.0012	1
.2	3.2532	3.2530	.0002	0	3.2532	0	0
.3	3.2611	3.2590	.0021	4	3.2597	.0014	2
.4	3.2516	3.2497	.0019	4	3.2506	.0010	1
.5	3.2282	3.2251	.0031	10	3.2259	.0023	5
.6	3.1807	3.1851	-.0044	19	3.1856	-.0049	24
.7	3.1266	3.1299	-.0033	11	3.1297	-.0031	10
.8	3.0594	3.0594	0	0	3.0582	.0012	1
.9	2.9759	2.9735	.0024	6	2.9711	.0048	23
1.0		2.8724			2.8684		
				.000057			
							.000070

Using the ordinary least squares test for accuracy the probable error in any velocity calculated by the least squares method is

$$r = .6745 \sqrt{\frac{.000057}{10 - 3}} = .0019$$

The probable error in a velocity calculated by the formula derived by method of moments is

$$r = .6745 \sqrt{\frac{.000070}{10 - 3}} = .0021$$

The difference in the degree of accuracy is then so slight as to be negligible even when the test of accuracy is the one derived for least squares where the function minimized was the sum of

the squares of the residuals at the points chosen while the function minimized in the method of moments was the sum of the squares of all the residuals all along the curve. It seems then evident that even for the parabolic curve, where the method of least squares is most useful, in a curve where the final form required must be referred to a particular axis, and using the ordinary least squares test for accuracy, the method of moments gives results sensibly as good as least squares with less labor.

When the problem does not include transfer of moments from one axis to another the work is much less. As an example let it be required to fit a second degree parabola to the following data which represent a cooling curve for water when surrounded by a freezing mixture. The x's are the time in minutes and the y's are observed Centigrade temperatures. x's are results of transferring the y-axis to $x = 10$ the mid-point of the range. Quadrature was by Simpson's rule.

$$\int z \, dx = \frac{1}{3}h \left[z_0 + z_{2p} + 2(z_2 + z_4 + \dots + z_{2p-2}) + 4(z_1 + z_3 + \dots + z_{2p-1}) \right]$$

(See section on quadrature, pages 35-47.

For this problem $h = 1$.

The work is tabulated as described on page 51, except that the numerical factors are 1, 4 and 2 instead of 1, 3 and 2. It depends of course on the number of observations what quadrature

formula can be used.

Assume

$$y = y_0 \left[e_0 + e_1 \frac{x}{1} + e_2 \left(\frac{x}{1} \right)^2 \right]$$

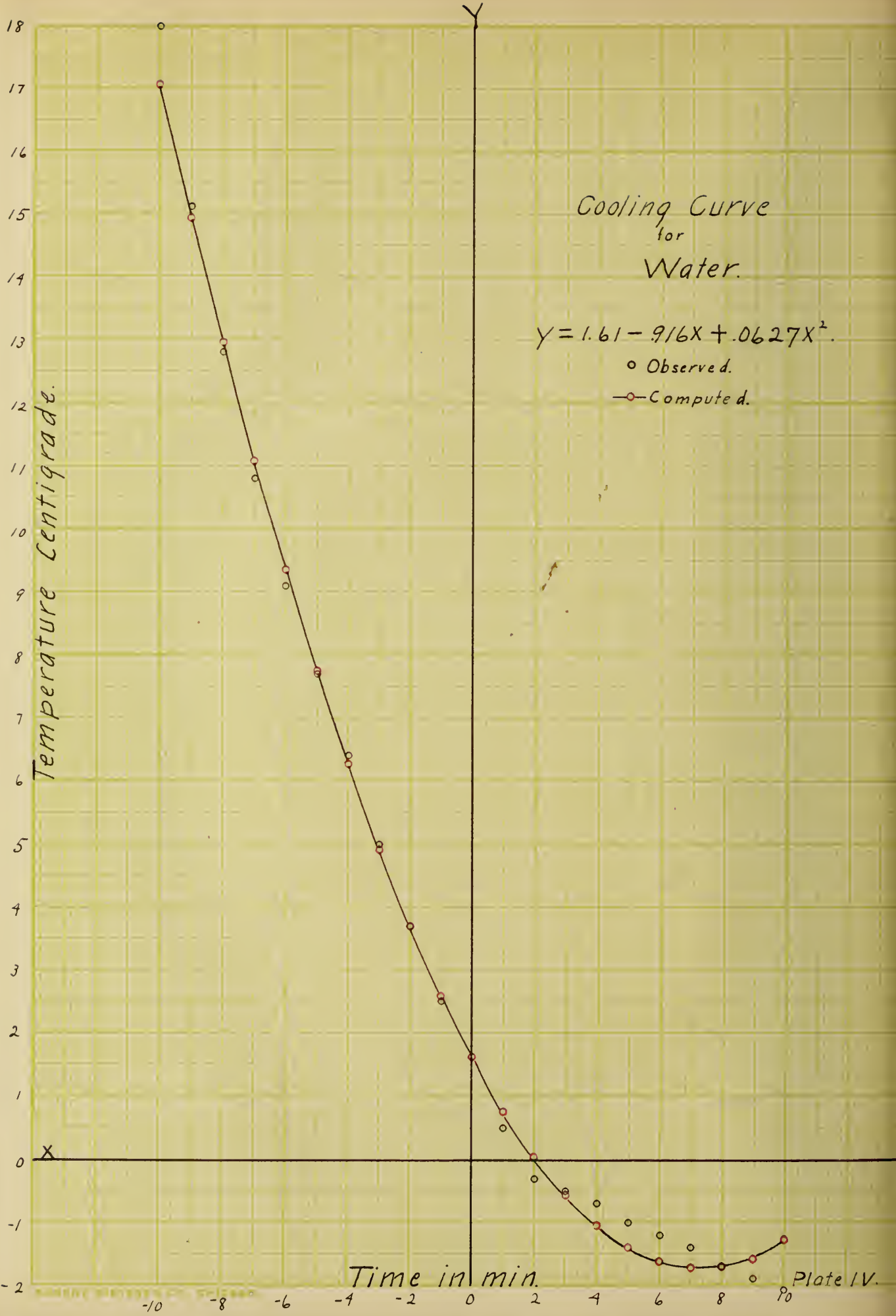
x	x'	y	S(y)	x'y	S(x'y)	x' ² y	S(x' ² y)
0	-10	18.0	18.0	-180.0	- 180.0	1800.0	1800.0
1	- 9	15.1	60.4	-135.9	- 543.6	1223.1	4892.4
2	- 8	12.8	25.6	-102.4	- 204.8	819.2	1638.4
3	- 7	10.8	43.2	- 75.6	- 302.4	529.2	2116.8
4	- 6	9.1	18.2	- 54.6	- 199.2	327.6	655.2
5	- 5	7.7	30.8	- 38.5	- 154.0	192.5	770.0
6	- 4	6.4	12.8	- 25.6	- 51.2	102.4	204.8
7	- 3	5.0	20.0	- 15.0	- 60.0	45.0	180.0
8	- 2	3.7	7.4	- 7.4	- 14.8	14.8	29.6
9	- 1	2.5	10.0	- 2.5	- 10.0	2.5	10.0
10	0	1.45	2.9	0	0	0	0
11	1	.5	2.0	.5	2.0	.5	2.0
12	2	.3	.6	.6	1.2	1.2	2.4
13	3	.5	2.0	1.5	6.0	4.5	18.5
14	4	.7	1.4	2.8	5.6	11.2	22.4
15	5	1.0	4.0	5.0	20.0	25.0	100.0
16	6	1.2	2.4	7.2	14.4	43.2	86.4
17	7	1.4	5.6	9.8	39.2	68.6	274.4
18	8	1.7	3.4	13.6	27.2	108.8	217.6
19	9	1.9	7.6	17.1	68.4	153.9	615.6
20	10	2.1	<u>2.1</u>	- 21.0	<u>- 21.0</u>	-210.0	<u>-210.0</u>
			222.2		-1831.0		10752.4

Hence

$$AM_0 = A = \frac{222.2}{3} = 74.07$$

$$AM_1 = - \frac{1831.0}{3} = -610.33$$

$$AM_2 = \frac{10752.4}{3} = 3584.13$$



$$L_0 = 1$$

$$L_1 = \frac{M_1}{1} = \frac{-610.33}{74.07 \times 10} = -.8240$$

$$L_2 = \frac{M_2}{1^2} = \frac{3584.13}{74.07 \times 100} = .4839$$

$$e_0 = \frac{3(3L_0 - 5L_2)}{4} = .4354$$

$$e_1 = 3L_1 = -2.472$$

$$e_2 = \frac{15(3L_2 - L_0)}{4} = 1.694$$

$$y_0 = \frac{A}{2 \cdot 1} = \frac{74.07}{20} = 3.704$$

Hence equation is

$$y = 3.704(.4354 - .2472 x + .01694 x^2)$$

$$= 1.61 - .916 x + .0627 x^2$$

The following table shows the differences between the observed values of temperatures and those which would follow the law expressed in this equation.

x	x'	Observed y	Computed y	Difference v	v ²
0	-10	18.0	17.04	.96	.9216
1	-9	15.1	14.93	.17	.0289
2	-8	12.8	12.95	-.15	.0225
3	-7	10.8	11.09	-.29	.0841
4	-6	9.1	9.36	-.26	.0676
5	-5	7.7	7.76	-.06	.0036
6	-4	6.4	6.28	.12	.0144
7	-3	5.0	4.92	.08	.0064
8	-2	3.7	3.69	.01	.0001

x	x'	Observed y	Computed y	Difference v	v ²
9	- 1	2.5	2.59	-.09	.0081
10	0	1.45	1.61	-.16	.0256
11	1	.5	.76	-.26	.0676
12	2	- .3	.03	-.33	.1089
13	3	- .5	- .57	.07	.0049
14	4	- .7	-1.05	.35	.1225
15	5	-1.0	-1.40	.40	.1600
16	6	-1.2	-1.63	.43	.1849
17	7	-1.4	-1.73	.33	.1089
18	8	-1.7	-1.70	0	0
19	9	-1.9	-1.56	.34	.1156
20	10	-2.1	-1.28	.82	<u>.6724</u>
					2.7286

The probable error in any computed y is then

$$r = .6745 \sqrt{\frac{2.7286}{21 - 3}} = .26$$

The inaccuracy of fit at the end points will be noted. This is to be expected as a result of all the quadrature formulas giving less weight to the end points and is one of the weak places of the method. Omitting the end points reduces the probable error to .17. The average of the absolute values of y is 4.94 of which the probable error is about 5 %.

EXPONENTIAL FUNCTION.

Next in order of usefulness to the parabolic function is the exponential function

$$y = e^{f(x)}$$

This is most easily dealt with by passing to logarithms,

$$Y = \log_e y = f(x)$$

and fitting this form of curve to logarithms of data. $f(x)$ may of course be any function of x and the form to be assumed would depend on the nature of the curve obtained by plotting a set of points with abscissae x and ordinates $Y = \log_e y$.

If the form of this curve is parabolic, then the method developed in the last section will give a curve fitting

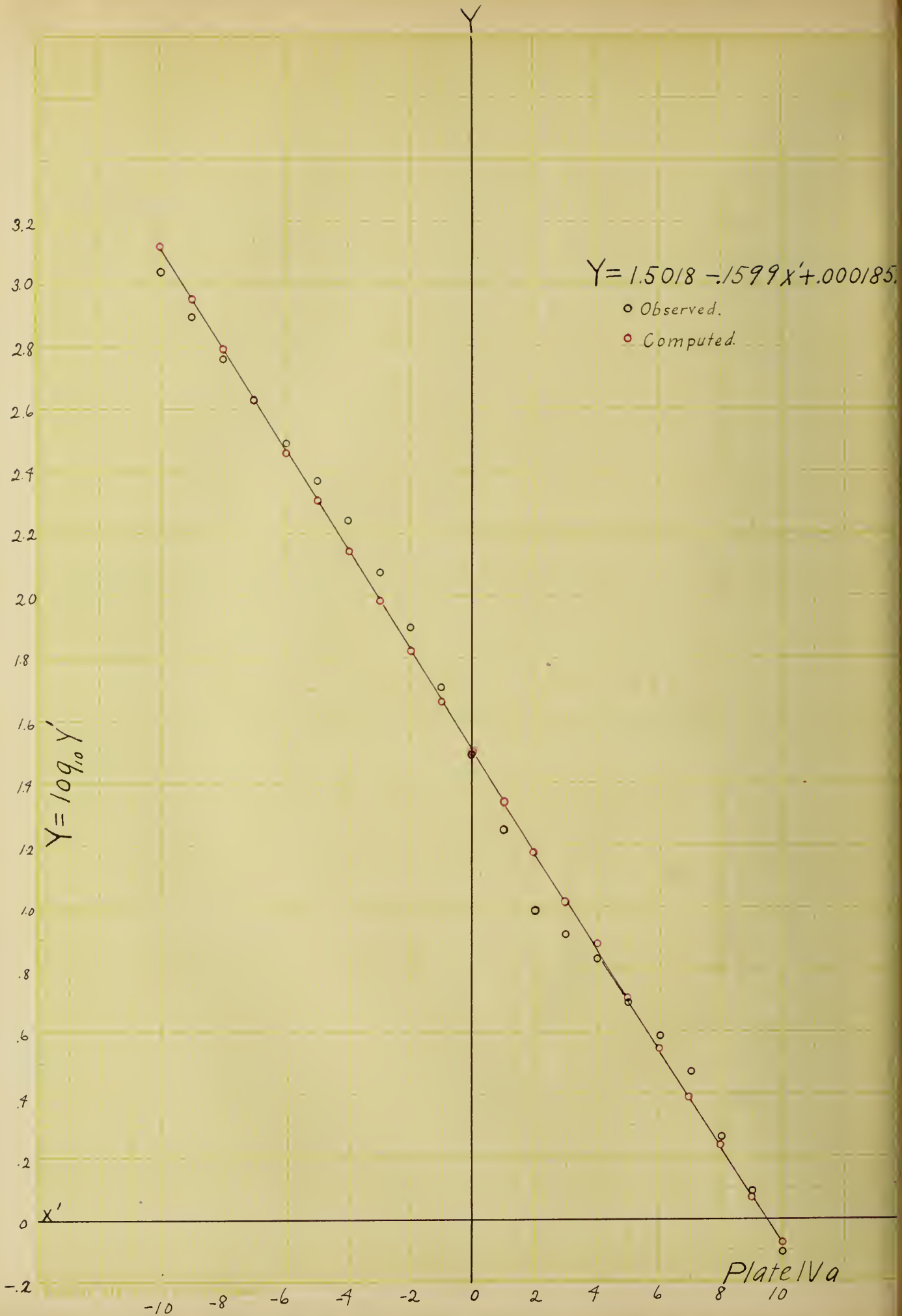
$$Y = f(x)$$

and hence

$$y = e^{f(x)}$$

Let it be required to fit a curve of form

$$y = e^{a + bx + cx^2}$$



to the data on the cooling of water discussed in the last problem. The first difficulty is that there is no logarithm of a negative number. Hence the first step is a transfer of axis moving the x-axis down three units, by the substitution $y' = y + 3$. The y-axis will also be taken in the middle of the range by the substitution $x' = x - 10$. In the following table x is time in minutes, y observed temperature in degrees Centigrade, x' and y' as given above, $Y = \log_e y'$ and other columns are for quadrature. Simpson's rule was used in quadrature. (See pages 35-47.) $h = 1$

The assumed function is

$$Y = a + bx' + cx'^2 = y_0 \left[e_0 + e_1 \frac{x'}{1} + e_2 \left(\frac{x'}{1} \right)^2 \right]$$

x	y	x'	y'	Y	S(Y)	x'Y	S(x'Y)	x' ² Y	S(x' ² Y)
0	18.0	-10	21.0	3.0445	3.0445	-30.4450	-30.4450	304.4500	304.4500
1	15.1	-9	18.1	2.8959	11.5836	-26.0631	-104.2524	234.5679	938.2716
2	12.8	-8	15.8	2.7600	5.5200	-22.0800	-44.1600	176.6400	553.2800
3	10.8	-7	13.8	2.6246	10.4984	-18.3722	-73.4888	128.6054	514.4216
4	9.1	-6	12.1	2.4932	4.9864	-14.9592	-29.9184	89.7552	179.5104
5	7.7	-5	10.7	2.3702	9.4808	-11.8510	-47.4040	59.2550	257.0200
6	6.4	-4	9.4	2.2407	4.4814	-8.9628	-17.9256	35.8512	71.7024
7	5.0	-3	8.0	2.0794	8.3176	-6.2382	-24.9528	18.7146	74.8584
8	3.7	-2	6.7	1.9021	3.8042	-3.8042	-7.6084	7.6084	15.2168
9	2.5	-1	5.5	1.7047	6.8188	-1.7047	-6.8188	1.7047	6.8188
10	1.45	0	4.45	1.4929	2.9858	0	0	0	0
11	.5	1	3.5	1.2527	5.0108	1.2527	5.0108	1.2527	5.0108
12	-.3	2	2.7	.9932	1.9864	1.9864	3.9728	3.9728	7.9456
13	-.5	3	2.5	.9163	3.6652	2.7489	10.9956	8.2467	32.9868
14	-.7	4	2.3	.8329	1.6658	3.3316	6.6632	13.3264	26.6528
15	-1.0	5	2.0	.6932	2.7728	3.4660	13.8640	17.3300	69.3200
16	-1.2	6	1.8	.5878	1.1756	3.5268	7.0536	21.1608	42.3216
17	-1.4	7	1.6	.4700	1.8800	3.2900	13.1600	23.0300	92.1200

x	y	x'	y'	Y	S(Y)	x'y	S(x'y)	x' ² Y	S(x' ² Y)
18	-1.7	8	1.3	.2623	.5246	2.0984	4.1968	16.7872	33.5744
19	-1.9	9	1.1	.0953	.3812	.8577	3.4208	7.7195	30.8772
20	-2.1	10	.9	-.1054	<u>-.1054</u>	-1.0540	<u>-1.0540</u>	-10.5400	<u>-10.5400</u>
					90.4785			-388.0282	3025.8192
							<u>68.3376</u>		
									-319.6906

$$AM_0 = \frac{1}{3} S(Y) = 30.1595 = A$$

$$AM_1 = \frac{1}{3} S(x'Y) = -106.5635$$

$$AM_2 = \frac{1}{3} S(x'^2Y) = 1008.6064$$

$$2 \ 1 = 20$$

$$1 = 10$$

$$y_0 = \frac{A}{2 \ 1} = \frac{30.1595}{20} = 1.50798$$

$$L_0 = \frac{M}{1} = 1$$

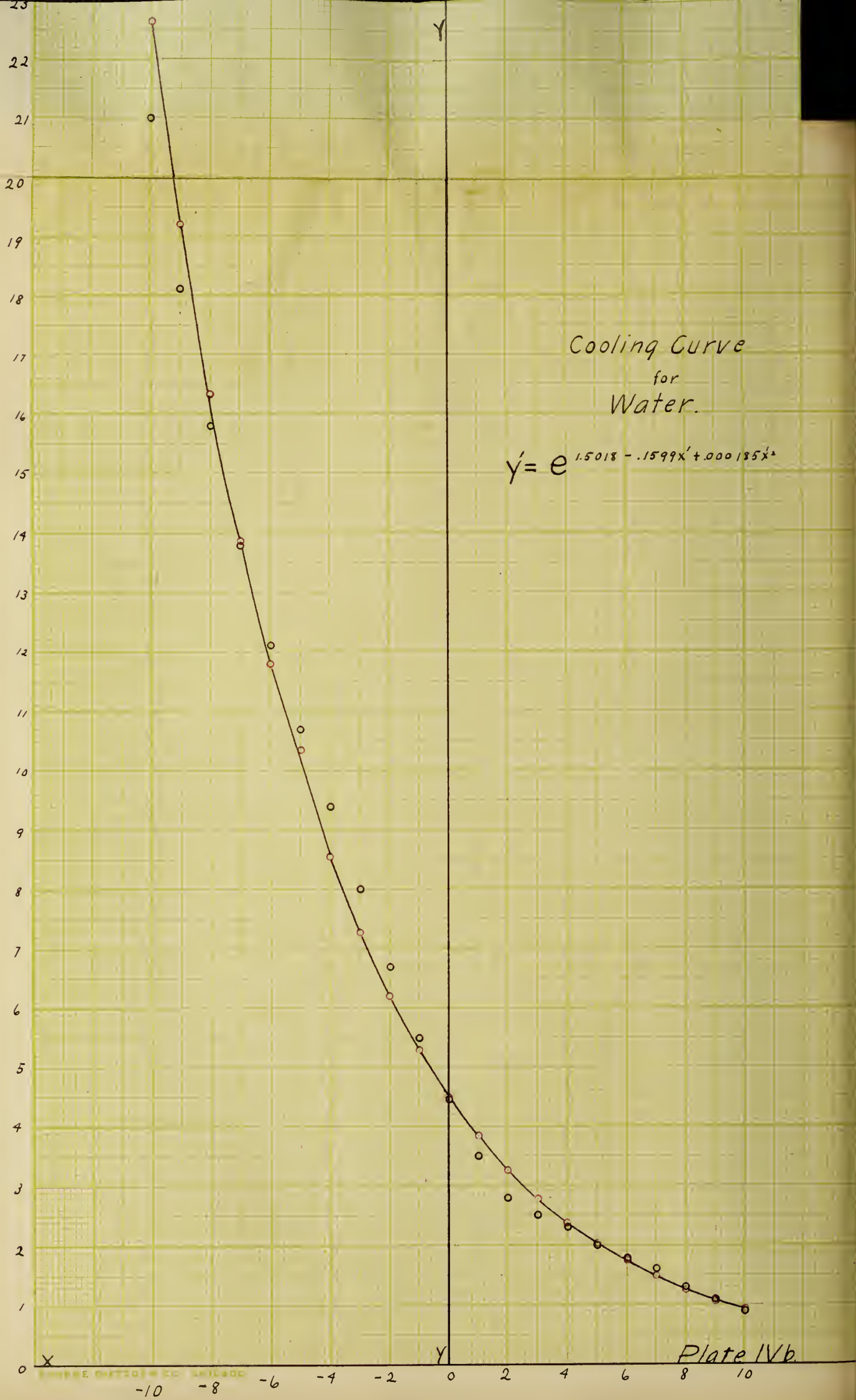
$$L_1 = \frac{M}{1} = -.353333$$

$$L_2 = \frac{M}{1} = .334424$$

$$e_0 = \frac{3}{4} (3 L_0 - 5 L_2) = .995910$$

$$e_1 = 3 L_1 = 1.059999$$

$$e_2 = \frac{15}{4} (3 L_2 - L_0) = .012270$$



Substituting in the assumed function

$$Y = 1.50798 \left(.995910 - .1059999 x' + .0001227 x'^2 \right)$$

$$= 1.5018 - .1599 x' + .000185 x^2$$

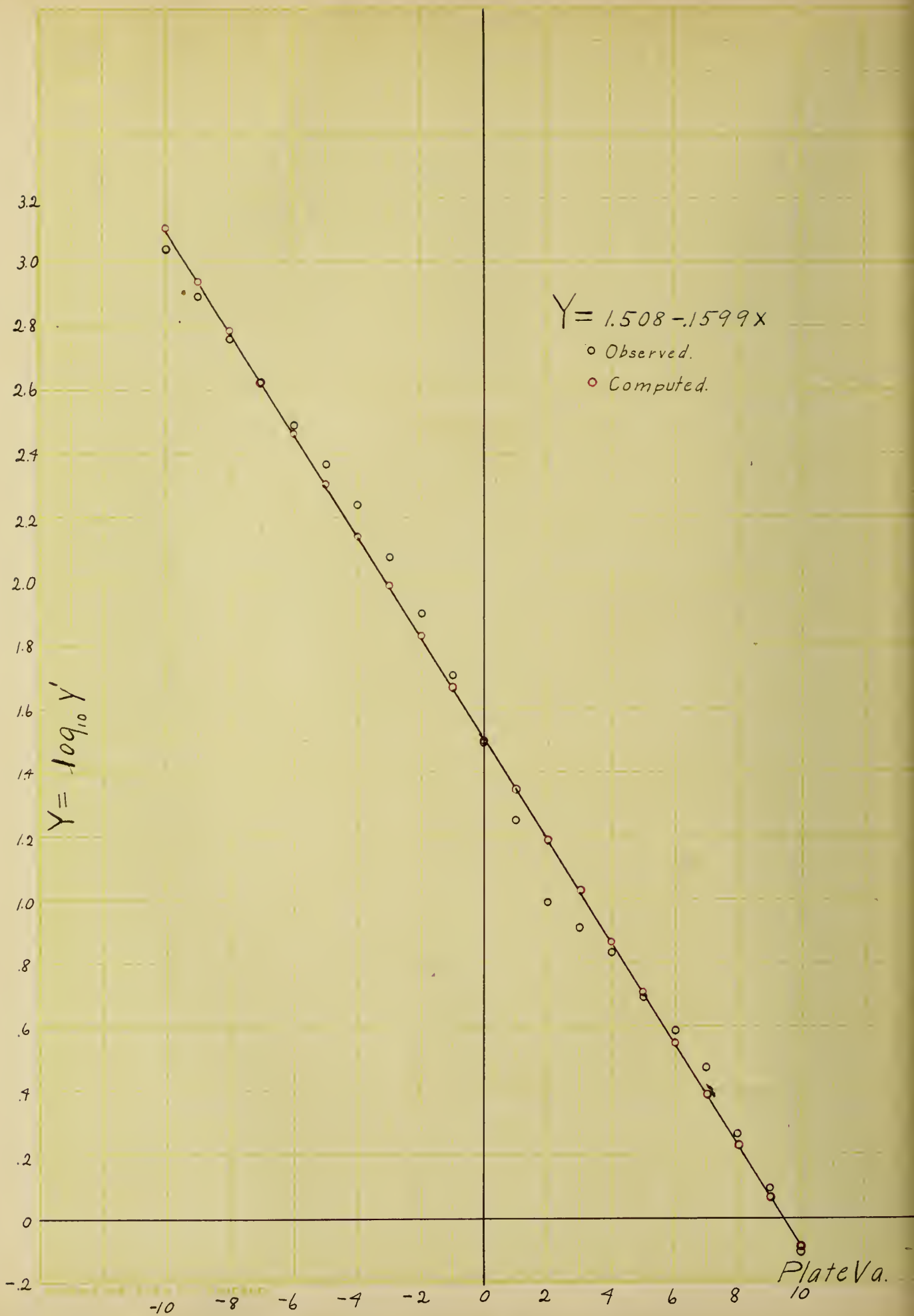
$$y' = e^{1.5018 - .1599 x' + .000185 x^2}$$

$$y = y' - 3$$

$$y = e^{1.5018 - .1599 x' + .000185 x^2} - 3$$

The following table shows the observed and computed values of y' with the differences between them.

x	x'	Ob- served Y	Com- puted Y	Differ- ence	Ob- served y'	Com- puted y'	Differ- ence v	v ²
0	-10	3.0445	3.1191	-.0746	21.0	22.63	-1.63	2.6569
1	- 9	2.8959	2.9557	-.0598	18.1	19.20	-1.10	1.2100
2	- 8	2.7600	2.7927	-.0327	15.8	16.32	- .52	.2704
3	- 7	2.6246	2.6300	-.0054	13.8	13.88	- .08	.0064
4	- 6	2.4932	2.4677	.0255	12.1	11.80	.30	.0900
5	- 5	2.3702	2.3058	.0644	10.7	10.34	.36	.1296
6	- 4	2.2407	2.1443	.0964	9.4	8.54	.86	.7396
7	- 3	2.0794	1.9851	.0913	8.0	7.28	.72	.5184
8	- 2	1.9021	1.8223	.0798	6.7	6.20	.50	.2500
9	- 1	1.7047	1.6618	.0429	5.5	5.29	.23	.0529
10	0	1.4929	1.5018	-.0089	4.45	4.49	- .04	.0016
11	1	1.2527	1.3421	-.0894	3.5	3.84	- .34	.1156
12	2	.9932	1.1828	-.1896	2.7	3.26	- .56	.3136
13	3	.9163	1.0238	-.1075	2.5	2.78	- .28	.0784
14	4	.8329	.8652	-.0323	2.3	2.38	- .08	.0064
15	5	.6932	.7070	-.0138	2.0	2.03	- .03	.0009
16	6	.5878	.5492	.0386	1.8	1.73	.07	.0049
17	7	.4700	.3917	.0783	1.6	1.49	.11	.0121
18	8	.2623	.2346	.0277	1.3	1.27	.03	.0009



x	x'	Ob- served Y	Com- puted Y	Differ- ence	Ob- served y'	Com- puted y'	Differ- ence v	v ²
19	9	.0953	.0779	.0174	1.1	1.08	.02	.0004
20	10	-.1054	-.0785	-.0269	.9	..92	-.02	.0004
								<u>6.4594</u>

The probable error in any ordinate is then

$$r = .6745 \sqrt{\frac{6.4594}{21 - 3}}$$

$$= .39$$

and the curve is not so good a fit as the parabolic curve with three constants.

The smallness of the coefficient of x^2 suggests that an exponential curve with two constants might give a curve which would fit the data fairly well. The fit will not of course be perfect for the plotted curve $Y = f(x)$ is not a straight line. This solution will be really fitting the best straight line to this curve and then taking the exponential of both members of the equation expressing the result.

Assume

$$Y = y_0 \left[e_0 + e_{11} \frac{x}{1} \right]$$

Only two constants are to be determined so area and first moment are sufficient. These were computed on pages 63-64 and the following values found.

22
21
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0

X'

-10 -8 -6 -4 -2 0 2 4 6 8 10

Plate Vb

X'

$y' = e^{1.508 - .1599x'}$
○ Observed.
● Computed.

23

$$L_0 = 1$$

$$L_1 = -.353333$$

By the general formulas for e's of parabolic curve, page 47,

$$e_0 = L_0 = 1$$

$$e_1 = 3L_1 = -1.059999$$

$$y_0 = 1.50798$$

page 64,

$$l = 10$$

$$Y = 1.50798(1 - .1059999 x)$$

Plate V_a

$$= 1.508 - .1599 x$$

$$y' = e^{1.508 - .1599 x}$$

$$y = e^{1.508 - .1599 x} - 3$$

Comparison of the values of y computed by this formula with observed values in the following table shows that the exponential formula in two constants gives results which are sensibly coincident with those given by the exponential in three constants, while of course its derivation is much less laborious.

x	x'	Observed Y	Computed Y	Observed y'	Computed y'
0	-10	3.0445	3.1070	21.0	22.36
1	- 9	2.8959	2.9471	18.1	19.06
2	- 8	2.7600	2.7872	15.8	16.23
3	- 7	2.6246	2.6273	13.8	13.84
4	- 6	2.4932	2.4674	12.1	11.80
5	- 5	2.3702	2.3075	10.7	10.54
6	- 4	2.2407	2.1476	9.4	8.57
7	- 3	2.0794	1.9877	8.0	7.30
8	- 2	1.9021	1.8272	6.7	6.22
9	- 1	1.7047	1.6679	5.5	5.30
10	0	1.4929	1.5080	4.45	4.52
11	1	1.2527	1.3481	3.5	3.86
12	2	.9932	1.1882	2.7	3.28
13	3	.9163	1.0283	2.5	2.79
14	4	.8389	.8634	2.3	2.38
15	5	.6932	.7085	2.0	2.03
16	6	.5878	.5486	1.8	1.73
17	7	.4700	.3887	1.6	1.48
18	8	.2623	.2288	1.3	1.26
19	9	.0953	.0689	1.1	1.07
20	10	- .1054	- .0910	.9	.91

THE SINE CURVE.

A type of curve often useful when the data plots a periodic or repeating curve is

$$y = a \sin nx + b \cos nx$$

which is simply a sine curve with the y-axis shifted horizontally to a new origin. For if

$$y = a \sin nx$$

The transformation $y = y'$

$$x = x' + k$$

gives $y = a \sin (nx + nk)$

$$= a \sin nx \cos nk + a \cos nx \sin nk$$

$$= b \sin nx + c \cos nx$$

by collecting all constants into a single coefficient.

If a curve of this type is to be fitted, to determine the three constants, a , b and n will require the area and first ^{two} moments. Let $2l$ be the total range of observations

and take the origin at the mid-point of the range. Then the area

$$(1) \quad AM_0 = A = \int_{-1}^1 y \, dx = \frac{a}{n} \left[-\cos nx \right] + \frac{b}{n} \sin nx \Big|_{-1}^1$$

$$= \frac{2b}{n} \sin nl$$

$$(2) \quad AM_1 = \int_{-1}^1 xy \, dx = a \int_{-1}^1 x \sin nx \, dx + b \int_{-1}^1 x \cos nx \, dx$$

$$= -\frac{a}{n} x \cos nx + \frac{a}{n^2} \sin nx + \frac{b}{n} x \sin nx + \frac{b}{n^2} \cos nx \Big|_{-1}^1$$

$$= -\frac{2a}{n} \cos nl + \frac{2a}{n^2} \sin nl$$

$$(3) \quad AM_2 = a \int_{-1}^1 x^2 \sin nx \, dx + b \int_{-1}^1 x^2 \cos nx \, dx$$

$$= -\frac{a}{n} x^2 \cos nx + 2 \frac{a}{n^2} x \sin nx + \frac{2a}{n^3} \cos nx + \frac{b}{n} x^2 \sin nx$$

$$+ \frac{2b}{n^2} x \cos nx - \frac{2b}{n^3} \sin nx \Big|_{-1}^1$$

$$= \frac{2b}{n} \sin nl + \frac{4b}{n^2} \cos nl - \frac{4b}{n^3} \sin nl$$

Dividing (3) by l^2 and then by (1)

$$\frac{M_2}{M_0 l^2} = 1 + \frac{2}{nl} \cot nl - \frac{2}{n^2 l^2}$$

Let $nl = z$

$$\frac{1}{2} \left[\frac{Mz}{M_0 l^2} - 1 \right] = \frac{1}{z} \cot z - \frac{1}{z^2}$$

$$\text{Let } B = \frac{1}{2} \left[\frac{Mz}{M_0 l^2} - 1 \right]$$

$$\text{Then } B = \frac{1}{z} \cot z - \frac{1}{z^2}$$

$$(4) \quad \frac{1}{z} + Bz = \cot z$$

Equation (4) then must be solved for z and since l is known

$$n = \frac{z}{l}$$

From (I)

$$b = \frac{n A M_0}{2 \sin z} = \frac{A M_0 z}{2 l \sin z}$$

From (II)

$$\begin{aligned} a &= \frac{-A M_1}{\frac{2}{n} l \cos z - \frac{2}{n^2} \sin z} = \frac{-A M_1}{2 \left(\frac{1}{z} \cos z - \frac{1}{z^2} \sin z \right)} \\ &= \frac{-A M_1}{2 l^2 \sin z \left(\frac{1}{z} \cot z - \frac{1}{z^2} \right)} = \frac{-A M_1}{2 B l^2 \sin z} \end{aligned}$$

Let it be required to fit a curve of the form

$$y = a \sin nx + b \cos nx$$

to this data.

x	y	x	y
0	125.0	5	71.6
1	145.6	8	-72.8
3	138.2	10	-138.7

To apply the formula deduced in the last paragraph the origin must be taken in the middle of the range by the transformation

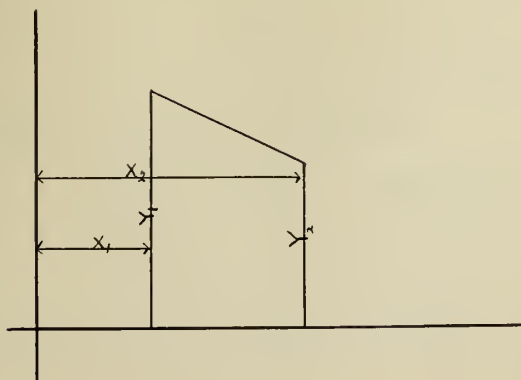
$$x' = x - 5$$

$$y' = y$$

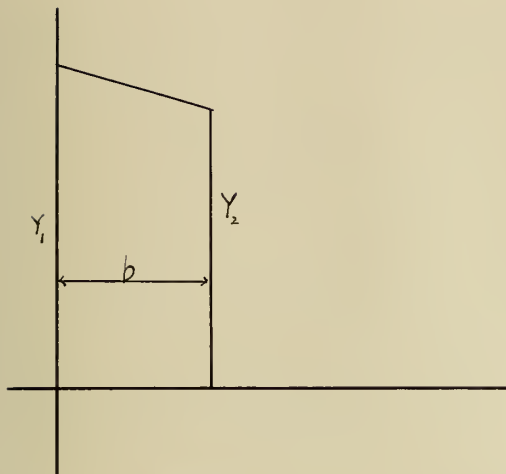
With this transformation the data becomes

x'	y	x'	y
-5	125.0	0	71.6
-4	145.6	3	-72.8
-2	138.2	5	-138.7

Since the ordinates are not equidistant the quadrature formulas of pp. 35-44 do not apply. The area is hence considered as a succession of trapezoids. The area of each is $1/2 (x_2 - x_1)(y_1 + y_2)$



For the first moments take axes as in figure.



$$x_1 = 0$$

Equation of line is

$$\frac{y - y_1}{x} = \frac{y_2 - y_1}{b}$$

$$y = \frac{y_2 - y_1}{b} x + y_1$$

$$A M_1 = \int_0^b xy \, dx = \frac{y_2 - y_1}{b} \frac{b^3}{3} + \frac{y_1 b^2}{2} = \frac{b^2(2y_2 + y_1)}{6}$$

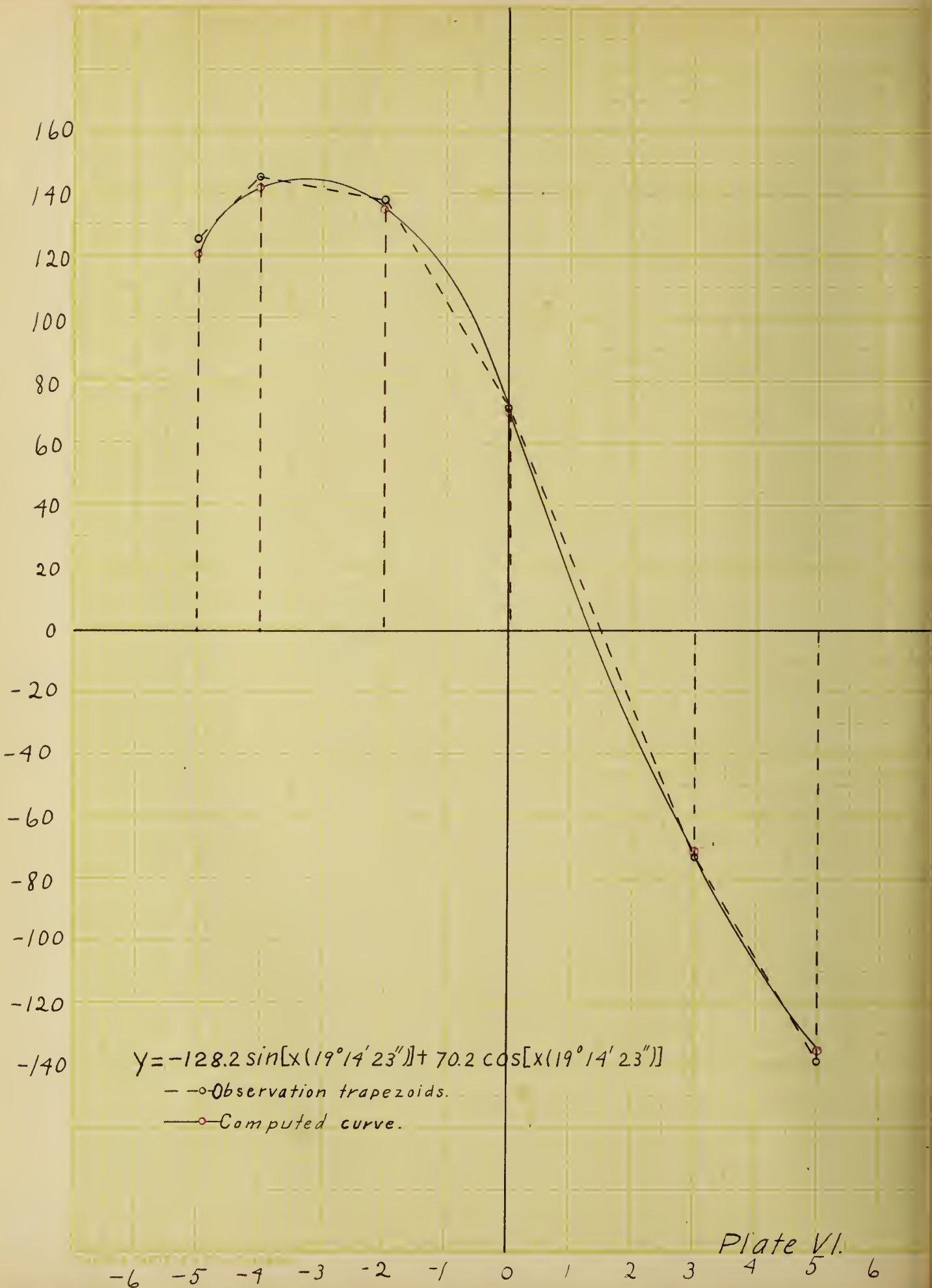
$$\bar{x} = \frac{A M_1}{A} = \frac{b^2(2y_2 + y_1)}{3b(y_1 + y_2)} = \frac{b(2y_2 + y_1)}{3(y_1 + y_2)}$$

This formula for first moment and centroid applies also to figures in which part of the area considered lies above and a part below the x-axis. For proper trapezoids the radius of gyration is nearly enough equal to $\bar{x}$ for purposes of this problem, but in case part of the area lies above and a part below the x-axis the second moment must be calculated for the two triangular parts separately.

In the table columns marked x and y are the data.

$x' = x - 5$ to bring the origin to the middle of the range

$b = x_2 - x_1$ is the base of the trapezoid, A is the area, $\frac{b C}{3 B}$



is distance from left hand boundary to centroid by formula on the previous page, $\bar{x} = x'_1 + \frac{bC}{3B}$.

x	y	x'	B $y_1 + y_2$	b	A	C $y_1 + 2y_2$	$\frac{bC}{3B}$	$\bar{x}$
0	125.0	-5	170.6	1	135.3	416.2	.51	-4.49
1	145.6	-4	283.8	2	283.8	422.0	.00	-3.01
3	138.2	-2	209.8	2	281.4	281.4	.89	-1.11
5	71.6	0	-1.2	3	-1.8	-74.0	61.70	61.67
8	-72.8	3	-211.5	2	-211.5	-350.2	1.10	4.10
10	-138.7	5						
415.6								

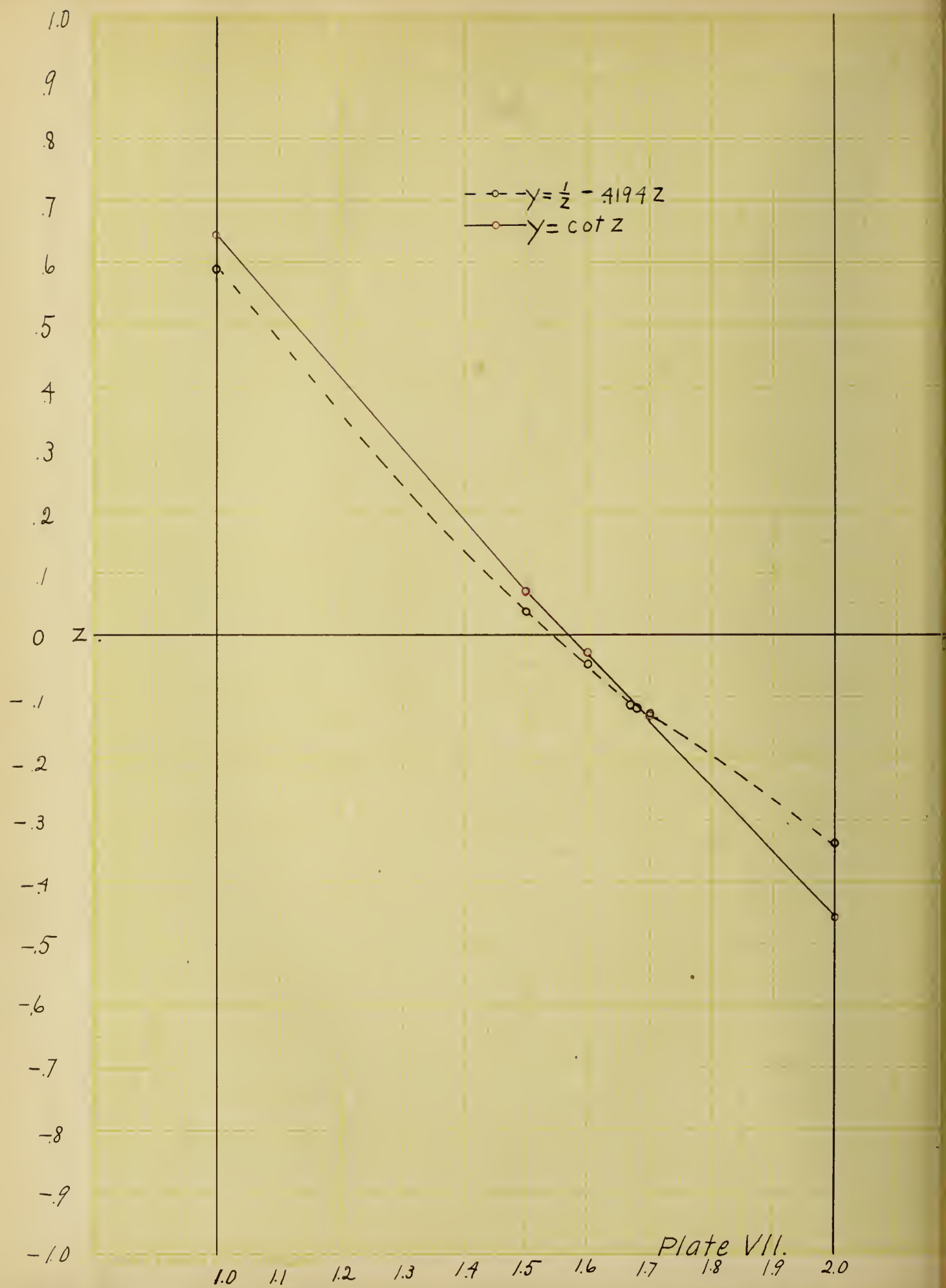
$AM_1 = \bar{x}A$	$\bar{x}^2$	$AM_2 = \bar{x}^2 A$
-607.50	20.16	2727.65
-854.24	9.06	2571.23
-232.88	1.23	258.05
-111.01	----	----
-867.15	16.81	-3555.32
		<u>- 328.10</u>
-2672.78		1673.51

Calculating the second moments of the two triangular portions of the fourth trapezoid separately and adding them gives -328.1. This sum is added in the last column in place of $\bar{x}^2 A$ which does not apply to figures of this kind.

By definition

$$B = \frac{1}{2} \left(\frac{AM_2}{AM_0^2} - 1 \right)$$

Substituting $1 = 5$ and values of AM_0 and AM_2 found on last page



$$B = -.4194$$

Hence the equation to be solved for z is

$$\frac{1}{z} - .4194 z = \cot. z$$

This equation will evidently be satisfied by simultaneous solutions of

$$(1) y = \frac{1}{z} - .4194 z \text{ and}$$

$$(2) y = \cot z \quad (z \text{ in radians of course.})$$

But each of these is a continuous curve and hence plotting as in accompanying figure, Plate VII, will help us to an approximate solution.

If $z = 1$, (1) gives $y = .5906$, (2) gives $y = .6421$

" $z = 2$, (1) " $y = -.3388$, (2) " $y = -.4576$

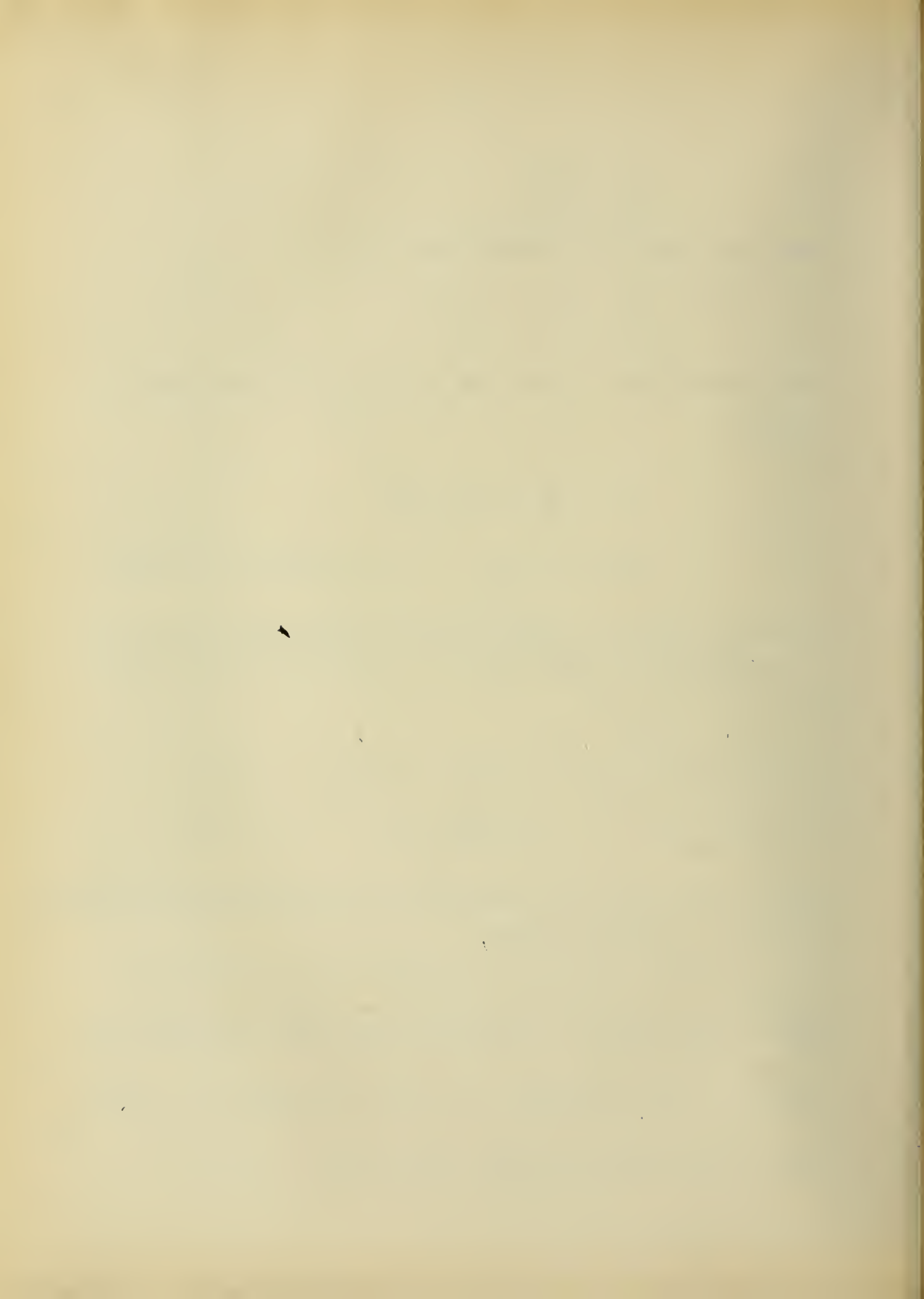
Plotting these shows that there must be an intersection and hence a value of z between 1 and 2.

Taking $z = 1.5$, (1) gives $y = .0375$, (2) gives $y = .0709$

Hence z lies between 1.5 and 2, and apparently much nearer 1.5.

Taking $z = 1.6$, (1) gives $y = -.0461$, (2) gives $y = -.0292$

Hence z lies between 1.6 and 2 nearer 1.6



Taking $z = 1.7$, (1) gives $y = -.1247$, (2) gives $y = -.1299$

Hence z lies between 1.6 and 1.7, nearer 1.7

Taking $z = 1.67$, (1) gives $y = -.1016$, (2) gives $y = -.0995$

Hence z lies between 1.67 and 1.7, nearer 1.67.

Taking $z = 1.68$, (1) gives $y = -.1094$, (2) gives $y = -.1096$.

Hence z lies between 1.67 and 1.68, nearer 1.68

Taking $z = 1.679$, (1) gives $y = -.1086$, (2) gives $y = -.1086$

and this is the value of z sought.

By hypothesis $z = n1 = 5n$

and hence $n = .3358 = 19^\circ 14' 23''$.

From formula I,

$$b = \frac{A M_0 z}{2 l \sin z} = \frac{415.6 \times 1.679}{10 \times .9942} = 70.2$$

From II,

$$a = \frac{-A M_1}{2 B l^2 \sin z} = \frac{-2672.78}{50 \times .4194 \times .9942}$$

$$= -128.2$$

Hence the equation of the curve is

$$y = -128.2 \sin [x(19^\circ 14' 23'')] + 70.2 \cos [x(19^\circ 14' 23'')]$$

The table shows the degree of accuracy of fit of the curve to data.



x	x'	Observed y	Computed y	Difference	% error
0	-5	125.0	119.9	5.1	4.1
1	-4	145.6	141.7	3.9	2.7
3	-2	138.2	134.7	3.5	2.6
5	0	71.6	70.2	1.4	1.9
8	3	- 72.8	- 70.9	1.9	2.6
10	5	-138.7	-135.0	3.7	2.7



THE HYPERBOLIC FORM.

Analogous to the parabolic curve in a power series of positive powers of x is the hyperbolic curve in negative powers of x . This curve is of the form

$$y = y_0 \left(a_0 + \frac{a_1}{x} + \frac{a_2}{x^2} + \dots \right)$$

This form of function is far less convenient to deal with from the stand point of integration than the parabolic form for the reason that any transfer to new axes alters the form of the function and not merely the constants involved in it. It is then studied here partly as an illustration of the fact that one can not always shift the axes about to suit his convenience and still retain the same form of function. Using so far as possible the same notation as in the preceding problems,

Let $2l$ = the range on the x -axis.

k = the lower limit of range.

Then $k + 2l$ = the upper limit of range.

$$\begin{aligned} AM_0 &= \int_k^{k+2l} y dx = y_0 \left[a_0 x + a_1 \log_e x - \frac{a_2}{x} + \dots \right]_k^{k+2l} \\ &= y_0 \left[a_0 \cdot 2l + a_1 \log_e \frac{k+2l}{k} + \frac{a_2 \cdot 2l}{k(k+2l)} + \dots \right] \end{aligned}$$

Handwritten text at the top of the page, possibly a title or header.

Handwritten text in the upper middle section, appearing to be a list or series of entries.

Handwritten text in the middle section, possibly a continuation of the list.

Handwritten text in the middle section, possibly a continuation of the list.

Handwritten text in the middle section, possibly a continuation of the list.

Handwritten text in the middle section, possibly a continuation of the list.

Handwritten text in the middle section, possibly a continuation of the list.

Handwritten text in the middle section, possibly a continuation of the list.

Handwritten text in the middle section, possibly a continuation of the list.

Handwritten text in the middle section, possibly a continuation of the list.

Handwritten text in the middle section, possibly a continuation of the list.

Handwritten text in the middle section, possibly a continuation of the list.

Handwritten text in the middle section, possibly a continuation of the list.

Handwritten text in the middle section, possibly a continuation of the list.

Handwritten text in the middle section, possibly a continuation of the list.

Handwritten text in the middle section, possibly a continuation of the list.

Handwritten text in the middle section, possibly a continuation of the list.

Handwritten text in the middle section, possibly a continuation of the list.

Handwritten text in the middle section, possibly a continuation of the list.

Handwritten text in the middle section, possibly a continuation of the list.

Handwritten text in the middle section, possibly a continuation of the list.

$$= 2l\gamma_0(a_0 + a_1 F + \frac{a_2}{K(K+2l)} \dots)$$

where

$$F = \frac{1}{2l} \cdot \log_e \frac{K+2l}{K}$$

$$\text{II } AM_1 = \int_K^{K+2l} xy \, dx = \gamma_0 \left[\frac{a_0}{2} x^2 + a_1 x + a_2 \log_e x + \dots \right]_K^{K+2l}$$

$$= \gamma_0 \left[\frac{a_0}{2} (4Kl + 4l^2) + a_1 2l + a_2 \log_e \frac{K+2l}{K} + \dots \right]$$

$$= 2l\gamma_0 [a_0(K+l) + a_1 + a_2 F + \dots]$$

$$\text{III } AM_2 = \int_K^{K+2l} x^2 y \, dx = \gamma_0 \left[\frac{a_0}{3} x^3 + \frac{a_1 x^2}{2} + a_2 x + \dots \right]_K^{K+2l}$$

$$= \gamma_0 \left[\frac{a_0}{3} (8l^3 + 12Kl^2 + 6lK^2) + \frac{a_1}{2} (4Kl + 4l^2) + a_2 2l + \dots \right]$$

$$= 2l\gamma_0 \left[\frac{a_0}{3} (4l^2 + 6Kl + 3K^2) + a_1 (K+l) + a_2 + \dots \right]$$

Since $A = 2l\gamma_0$

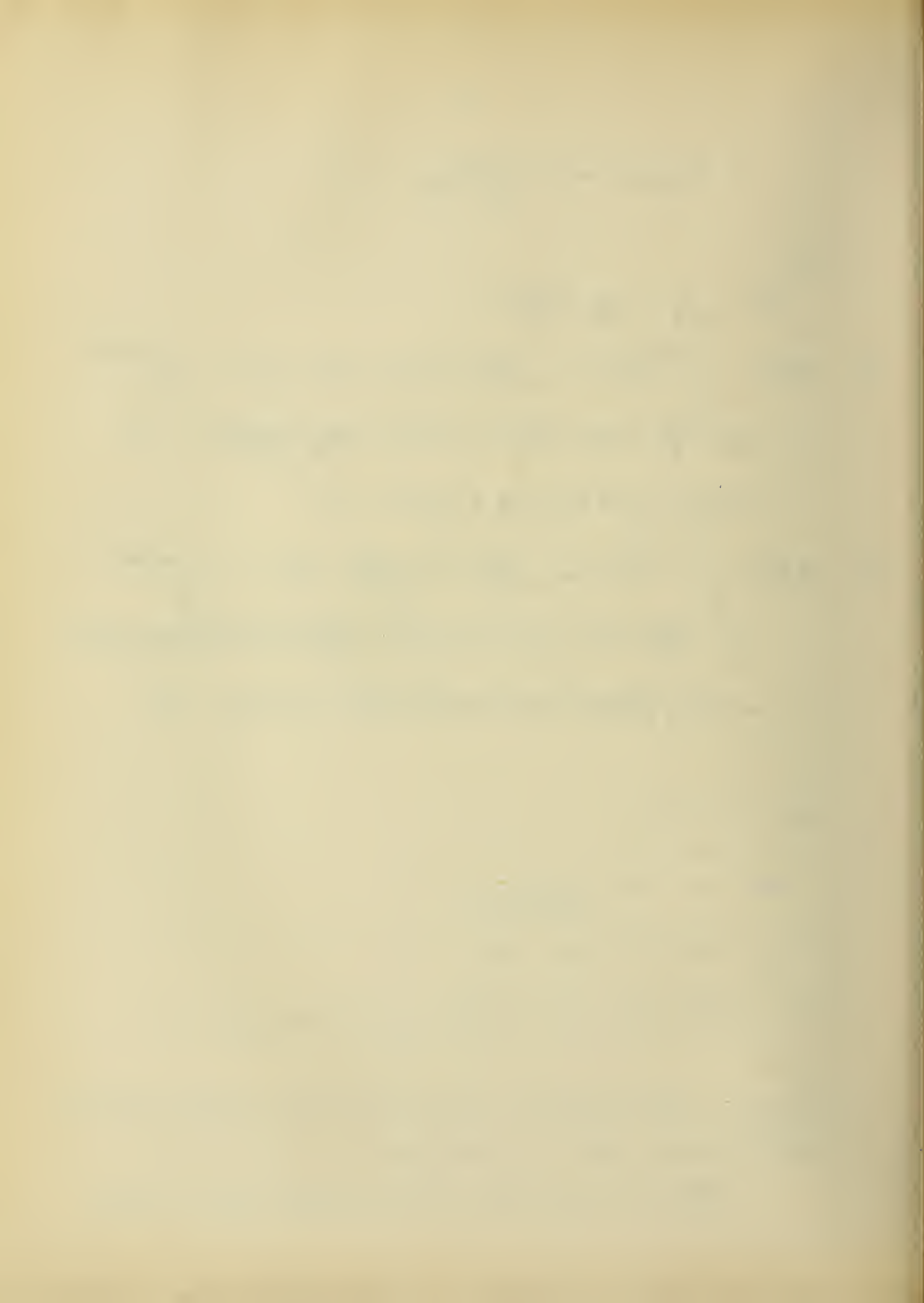
$$\text{I } M_0 = a_0 + a_1 F + \frac{a_2}{K(K+2l)} + \dots$$

$$\text{II } M_1 = a_0(K+l) + a_1 + a_2 F + \dots$$

$$\text{III } M_2 = \frac{a_0}{3} (4l^2 + 6Kl + 3K^2) + a_1 (K+l) + a_2 + \dots$$

It will be noticed that the complexity of the formulas increases rapidly with the number of constants.

A curve of this form in two constants will be fitted



to some data obtained from an experiment on Boyle's Law, for which a more perfect fit was desired than that given by the first term as ordinarily expressed

$$xy = k$$

or $y = k/x$

The assumed function is

$$y = y_0(a_0 + \frac{a_1}{x})$$

In the following table columns marked x and y give the data, x the volume, y the pressure. Since the ordinates are not equidistant, the commonly used quadrature formulas fail, and the areas will be considered a succession of trapezoids. b = the base of each trapezoid = $x_2 - x_1$, etc., as in tabulation of areas and moments for the sine curve.

x	y	b	B y_1+y_2	A	C y_1+2y_2	$\frac{bC}{3B}$	$\frac{-}{x}$	A M_1
29.0	118.5	.9	233.3	104.98	348.1	.45	29.45	3091.661
29.9	114.8	1.0	225.8	112.90	336.8	.50	30.40	3432.760
30.9	111.0	1.2	217.9	130.74	324.8	.60	31.50	4118.310
32.1	106.9	1.4	209.3	146.51	311.7	.70	32.80	4805.528
33.5	102.4	1.3	201.3	130.90	300.2	.65	34.15	4470.235
34.8	98.9	1.2	194.6	116.76	290.3	.60	35.40	4133.304
36.0	95.7	1.4	187.7	131.39	279.7	.70	36.70	4822.013
37.4	92.0	1.5	180.6	135.45	269.2	.75	38.15	5167.418
38.9	88.6	2.3	174.2	200.33	259.8	1.14	40.04	8021.213
41.2	85.6	2.5	164.2	205.25	242.8	1.23	42.43	8708.758
43.7	78.6	2.1	153.7	161.88	228.8	1.04	44.74	7242.511
45.8	75.1							
				1577.09				
								58013.711



$$A = 1577.09$$

$$M_0 = 1$$

$$M_1 = \frac{58013.711}{1577.09} = 36.7853$$

$$21 = 45.8 - 29 = 16.8$$

$$1 = 8.4$$

$$y_0 = 93.87$$

$$k = 29$$

$$k + 1 = 37.4$$

$$F = \frac{1}{21} \cdot \log_e \frac{k+21}{k} = .0272$$

Using formulas I and II to determine the constants

$$M_0 = a_0 + a_1 F$$

$$M_1 = a_0(k + 1) + a_1$$

$$(3) \quad a_1 = \frac{M_0 - a_0}{F} = M_1 - a_0(k + 1)$$

$$M_0 - a_0 = FM_1 - Fa_0(k + 1)$$

$$(4) \quad a_0 = \frac{FM_1 - M_0}{F(k + 1) - 1}$$

Substituting values in (4)

120

115

110

105

100

95

90

85

80

75

Pressure

Boyle's Law Data

$$y = 2.94 + \frac{3339.26}{x}$$

o Observed.

—o— Computed.

Volume

28

30

32

34

36

38

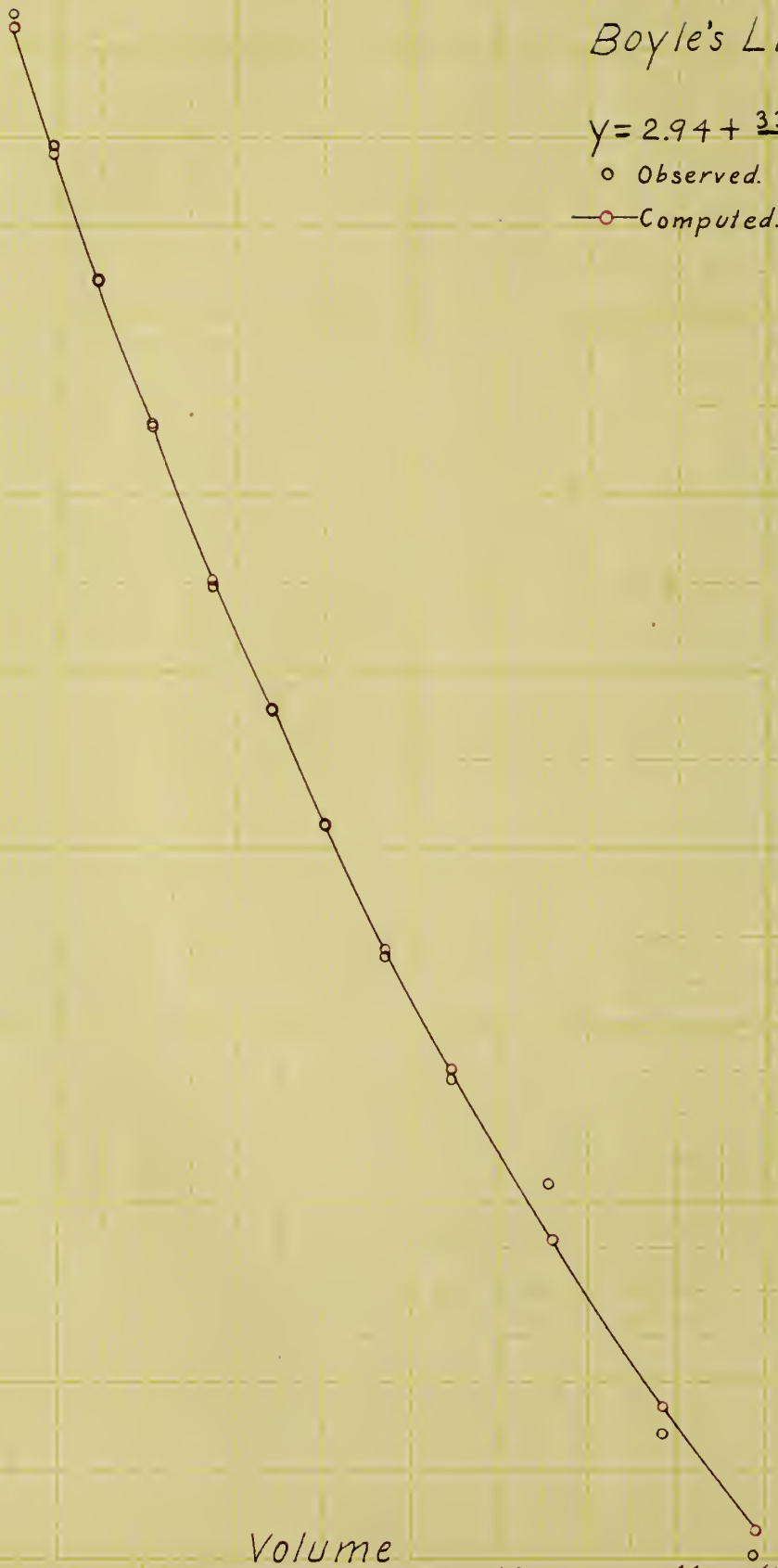
40

42

44

46

Plate VIII



$$a_0 = \frac{.0272 \times 36.7853 - 1}{.0272 \times 37.4 - 1} = .0324$$

Substituting in (3)

$$\begin{aligned} a_1 &= 36.7853 - .0324 \times 37.4 \\ &= 35.5732 \end{aligned}$$

Hence the curve is

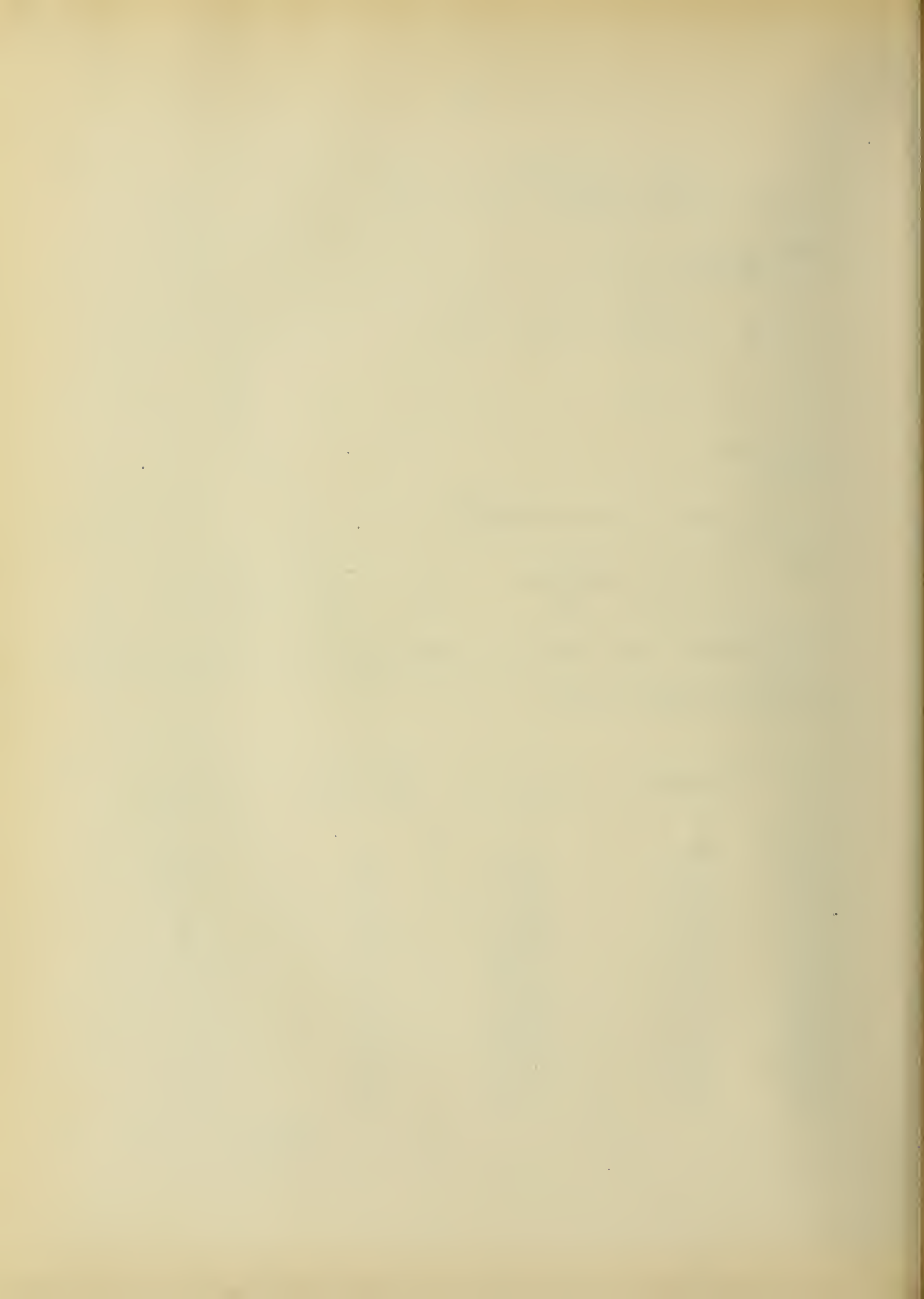
$$\begin{aligned} y &= 93.87 \left(.0324 + \frac{35.5732}{x} \right) \\ &= 2.94 + \frac{3339.26}{x} \end{aligned}$$

Plate VIII

The following table shows the closeness of fit of the theoretical and experimental curves.

x	Observed y	Computed y	Difference	% error
29.0	118.5	118.1	.4	.3
29.9	114.8	114.6	.2	.2
30.9	111.0	111.0	0	0
32.1	106.9	107.0	-.1	.1
33.5	102.4	102.6	-.2	.2
34.8	98.9	98.9	0	0
36.0	95.7	95.7	0	0
37.4	92.0	92.2	-.2	.2
38.9	88.6	88.8	-.2	.2
41.2	85.6	84.0	1.6	1.9
43.7	78.6	79.3	-.7	.9
45.8	75.1	75.8	-.7	.9

Average .4



The point whose coordinates are (41.2, 85.6) gives a % error $7 \frac{1}{2}$ times as large as the average of all the rest. This fact with the evident break made by this point in the observation curve leads one to suspect an error in the observation. Nevertheless the curve fits very well, and .4 % average error for a function in two constants shows the suitability of the form of function assumed.

In summing up the relative advantages and disadvantages of the methods of moments and least squares these seem to be the principal points.

1. The method of least squares will so determine the constants of an assumed function as to fit as well as possible the plotted points, but not necessarily the polygon formed by connecting these points.

2. More depends on the choice of form of function than on any great refinement of accuracy in determination of constants in a function of unsuitable form. Since the method of moments applies to a greater variety of forms, it has in this respect a decided advantage.

3. Even in the parabolic form to which the method of least squares is most easily applicable, the method of moments gives results sensibly as good with less labor, and with perfect quadrature formulas will give as good results.

For these reasons, it seems that for the practical



problems of the physicist and engineer the method of moments may be as useful as Prof. Pearson has already shown it in statistical work.



SPECIAL METHODS.

The advantage in certain cases of fitting a curve to the logarithms of data instead of directly to the data has been discussed and illustrated in a preceding section of this paper, and the use of logarithms as means of fitting curves of the form

$$y = kx^n$$

will now be considered.

Taking logarithms

$$\log y = \log k + n \log x$$

This is evidently a linear equation in two variables $\log x$ and $\log y$ with slope n and y -intercept $\log k$. If then on plotting logarithms of any set of data the graph is a straight line, the data can be fitted by a curve of the form

$$y = kx^n$$

and measuring the y -intercept gives $\log k$ and the tangent of the angle made with the x -axis is the value of n . As a mechanical device to avoid the labor of looking up logarithms, lo-



Discharge in cu. ft. per sec.

Plate IX.

Discharge in cu. ft. per sec.

Flow of water
thru a 4" orifice.
 $y = 435 X^{.525}$

X-axis.

Y-axis

Plate IX.

LOGARITHMIC SCALE, 10 INCH BASE

Head in ft.



logarithm paper may be used. It differs from other coordinate paper in that the length of the coordinate corresponding to the number printed in the margin is the logarithm of that number instead of the number itself. The spacing of the lines is thus exactly like that on a slide rule. The origin is marked (1,1) for the logarithms of these numbers are (0,0). If then the data be plotted on logarithm paper, the mathematical relations involved are exactly the same as if the logarithms of the data had been plotted on ordinary coordinate paper. In the accompanying figure, Plate IX, are plotted data of an experiment on flow of water through an orifice. x is the head in feet and y is the discharge in cubic feet per second.

x	y	x	y
.76	.377	1.41	.520
.89	.415	1.54	.547
1.02	.441	1.67	.565
1.15	.464	1.80	.589
1.28	.495		

Drawing a line which will fit these points as closely as possible, its y -intercept is found to be .435 and its slope (the ratio of its horizontal and vertical projections) to be .525. Hence the data can be fitted by

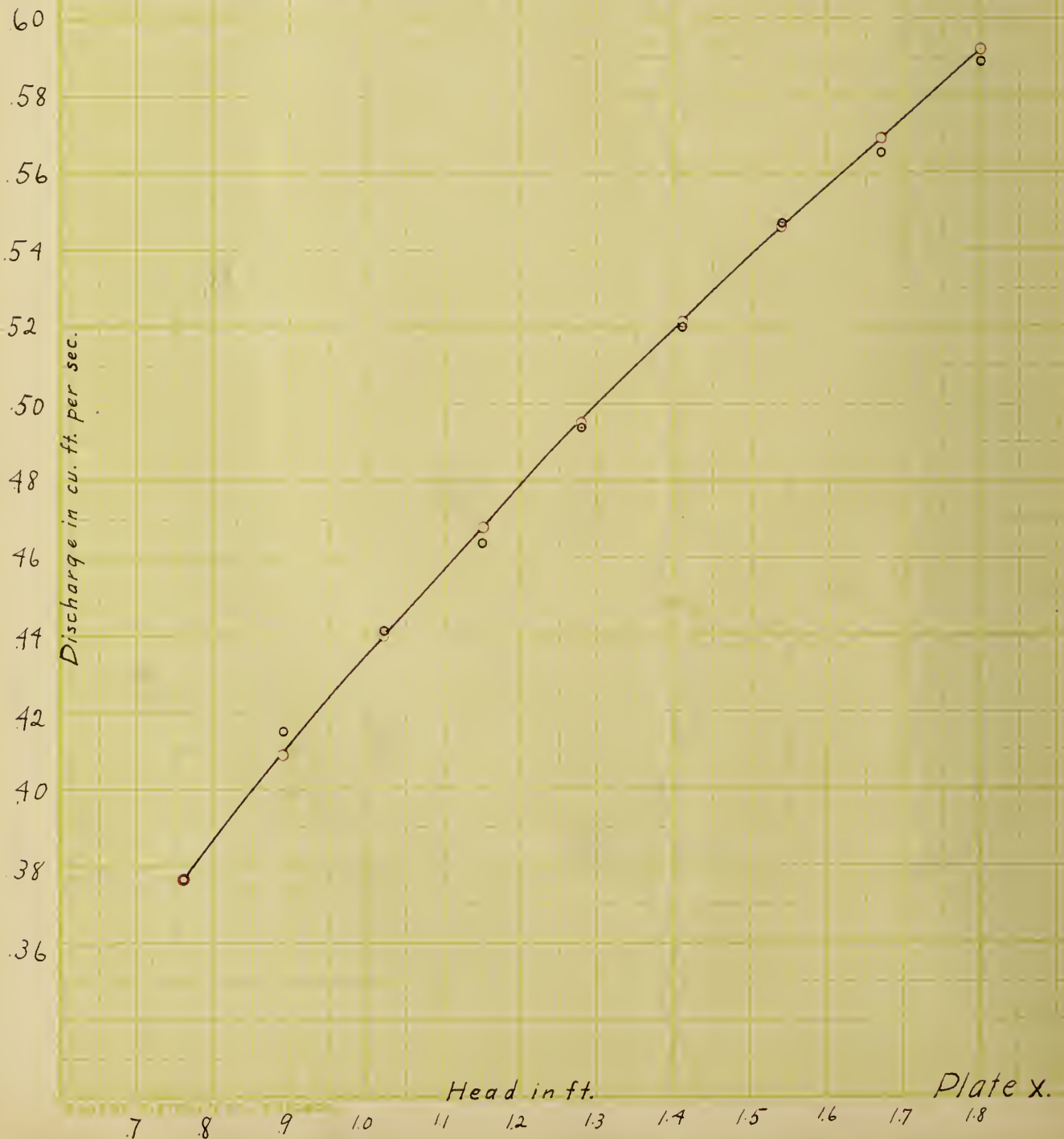
$$y = .435 x^{.525}$$

Flow of water
thru a 4" orifice.

$$y = 435 x^{.525}$$

○ Observed.

—○— Computed.



The table shows observed and computed values of y for each value of x , Plate IX.

x	Observed y	Computed y	Difference	% error
.76	.377	.377	0	0
.89	.415	.409	.006	1.4
1.02	.441	.440	.001	.2
1.15	.464	.468	-.004	.9
1.28	.494	.495	-.001	.2
1.41	.520	.521	-.001	.2
1.54	.547	.546	-.001	.2
1.67	.565	.569	-.004	.7
1.80	.589	.592	-.003	<u>.5</u>
Average				.5

As will be seen from the table the fit of the curve is very close, probably within the limit of accuracy of the observations. It will be seen that the formulas here derived differ slightly from the one commonly used.

$$y = c\sqrt{x}$$

where c is a constant for any particular orifice and value of x must be empirically determined and tabulated for each orifice over such range of variation in head as are likely to occur in connection with that orifice. Evidently for the experiment here worked out $c = .435 x^{.025}$

In the problems to which this method is applicable it forms a simple and valuable point of attack for determining



an unknown exponent and hence supplements the method of least squares and the method of moments at a point where both fail by leading to equations too complex to be solved. It is especially valuable in hydraulic formulas, though it applies of course wherever the function takes the form

$$y = kx^n$$

as in a period of a pendulum, relation of stress to deformation in concrete, etc.

Another use of logarithm paper in curve fitting applies to functions of the form

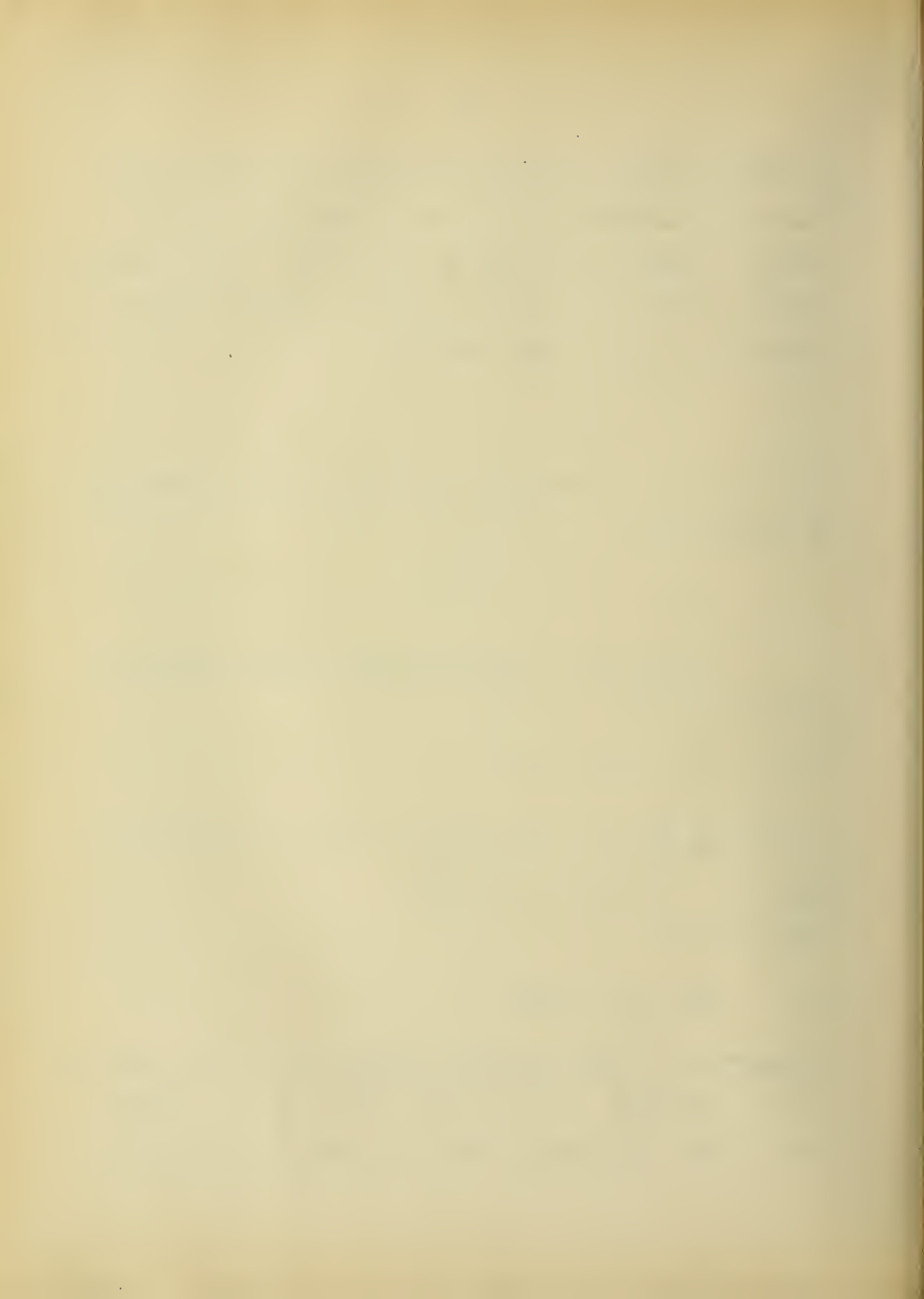
$$(1) \quad y - a = k(x - b)^n$$

a form which has four constants and has them so disposed as to give great flexibility to the curve.

Differentiating

$$(2) \quad \frac{dy}{dx} = nk(x - b)^{n-1}$$

a form which if b were eliminated would be exactly like that of the last paragraph. The value of b may be found as follows: Plot the data and connect consecutive points by straight lines.



Then if x_r and y_r are the coordinates of the r th point and x_{r+1} and y_{r+1} those of the $(r+1)$ th, the slope of the line joining these two points is $\frac{y_{r+1} - y_r}{x_{r+1} - x_r}$ which is a fair approximation to the value of $\frac{dy}{dx}$ at a point on the curve mid-way between the two. Call the abscissa of this point $x_{r+1/2}$. Evidently from (2)

$$\frac{dy}{dx} = 0$$

when $x = b$. If then corresponding values of $x_{r+1/2}$ and $\frac{dy}{dx}$ be plotted, the intersection with the x -axis will give the value of b . In some cases the curve will not intersect the x -axis within the range of observations. In such cases it must be prolonged until it does intersect. There is here of course a large chance for error but if b be taken incorrectly the fact will appear in the next step, so the process as a whole is self-checking. After b is determined the function is of the form previously studied,

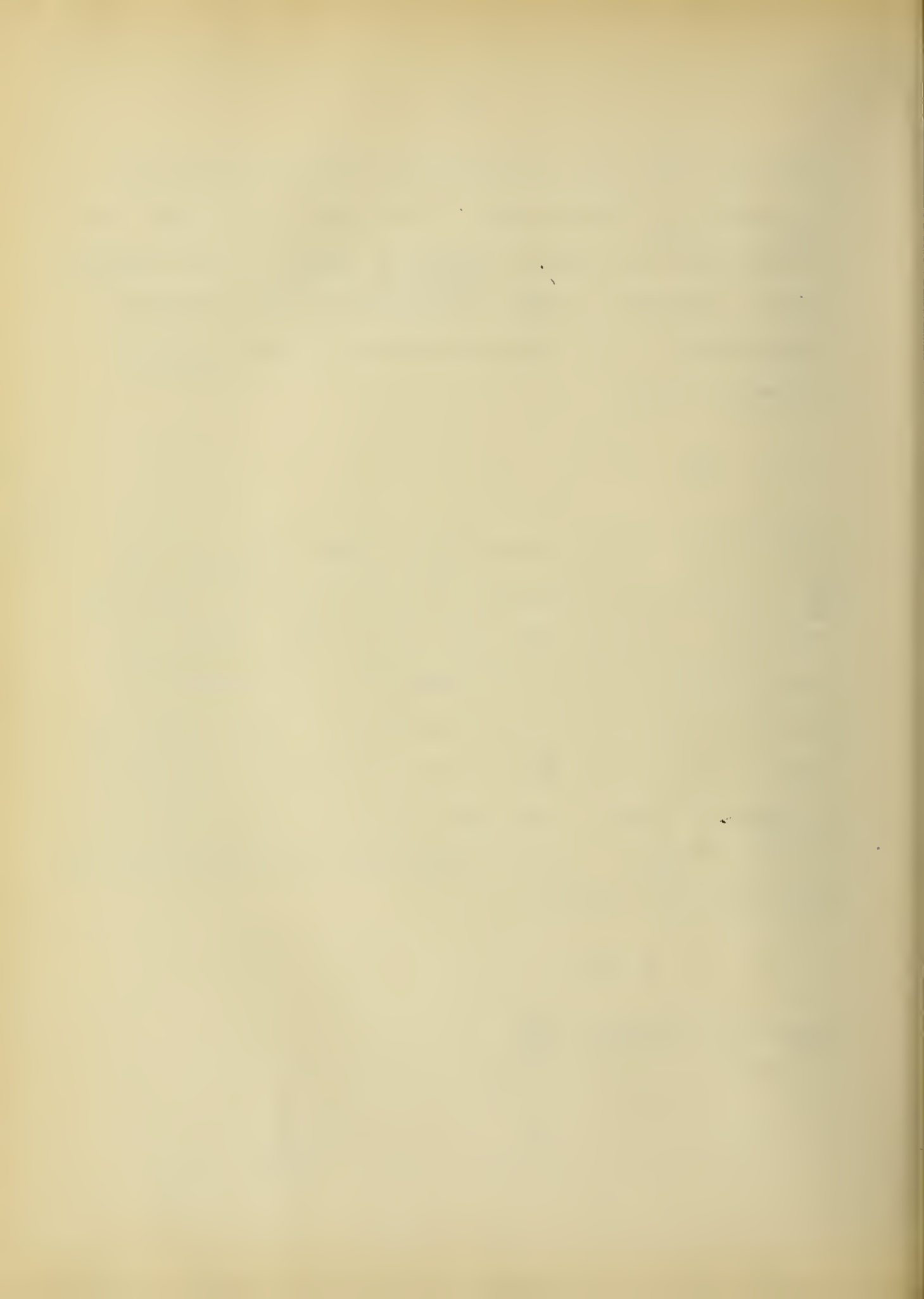
$$y = kx^n$$

where y is replaced by $\frac{dy}{dx}$

$$x \quad " \quad " \quad x - b$$

$$k \quad " \quad " \quad nk$$

$$n \quad " \quad " \quad n - 1$$

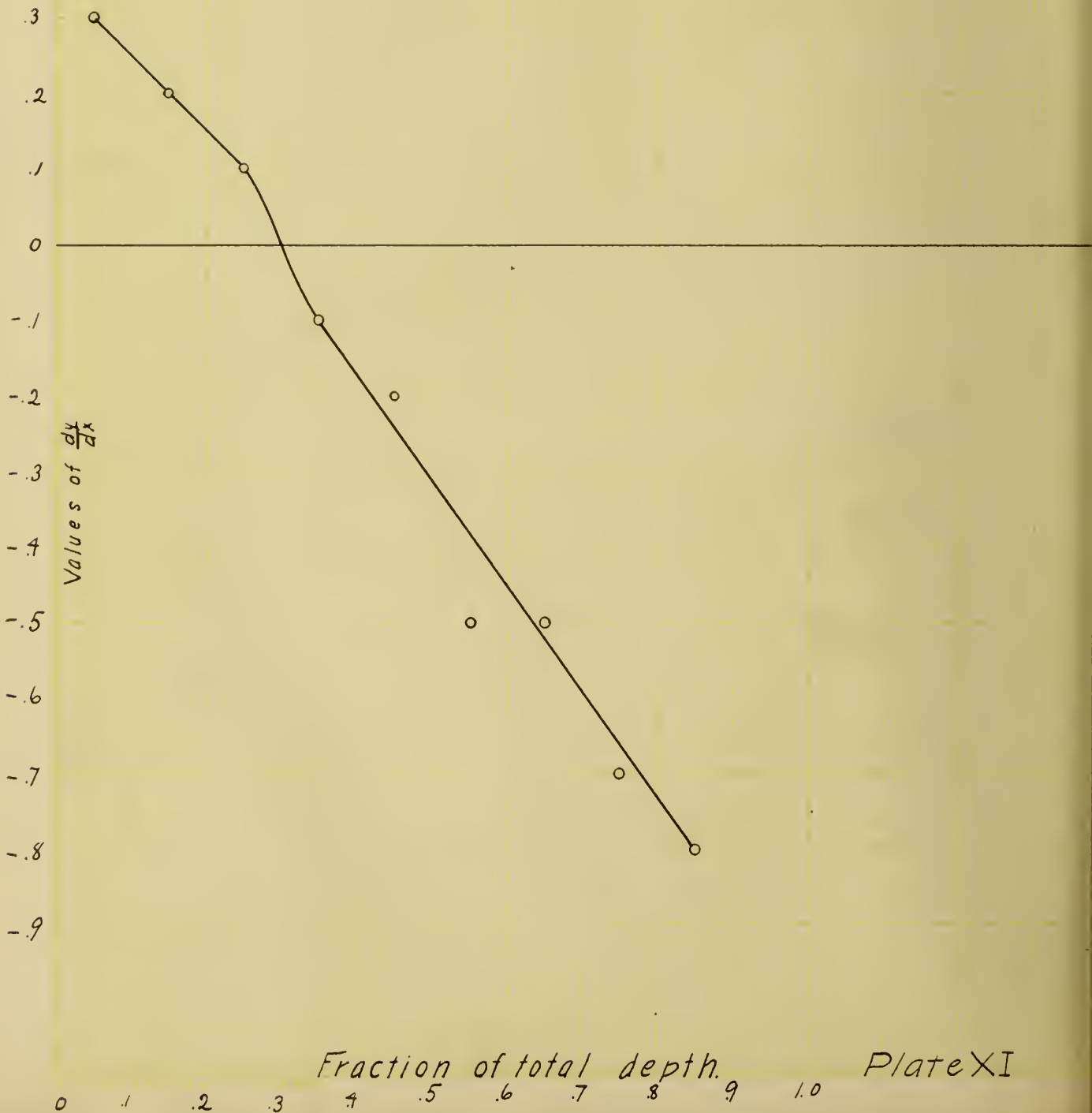


Corresponding values of $\frac{dy}{dx}$ and $x_{r+1/2} - b$ are then plotted on logarithm paper. If they plot a straight line b has been correctly chosen. If they form a regular curve b is in error. If the points are irregularly arranged this form of function is not well adapted to the data. In making successive trials to determine b it is useful to note that a value of b which is too large makes the curve convex on one side, and a value too small makes it convex on the other side. If two such curves have been found it is then known that the true value of b lies between the two which gave these curves and its value may be approximated by successive subdivisions of the interval. When a value of b is found which does make $\frac{dy}{dx}$ and $(x_{r+1/2} - b)$ plot a straight line on logarithm paper, the y-intercept of this line gives the value of nk and the slope the value of $(n - 1)$. These two equations may then be solved for k and n . All the constants of the original function are now determined except a , and it may now be found by substituting the data into this equation. This substitution forms a check on the accuracy of fit. A perfect fit would show the same value of a for each pair of values of x and y . This of course is not to be expected and an average value of a is taken and the error in any y is the difference between the value of a computed for that particular point and the average adopted for the formula.

Graph of
 X and $\frac{dy}{dx}$
 Mississippi River Data.

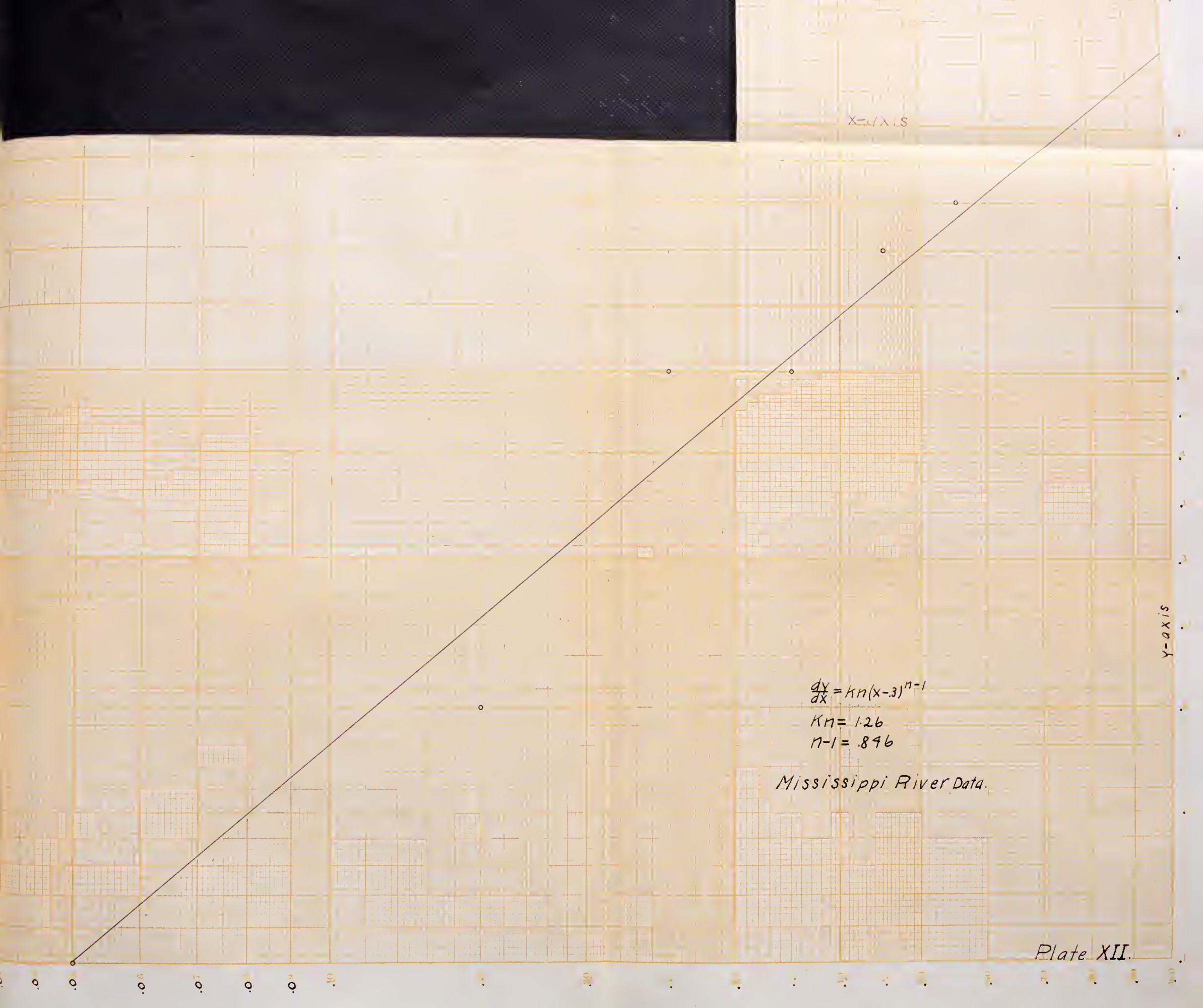
$$\frac{dy}{dx} = kn(x-b)^{n-1}$$

$$b = .3$$



From the graphical nature of the process, it is not adapted to data using a large number of significant figures. Hence in the problem treated, namely, the same one that was treated by least squares and by moments, the velocity of the of the Mississippi River at varying fractions of its total depth, the velocities are taken only to three significant figures.

In the table x and y are the data and they are tabulated on every other line so as to leave space for writing in the corresponding values of $x_{r+1/2}$ and $\frac{dy}{dx}$. The latter is the ratio of the y -increment to the x -increment in and since the x -increment is uniformly .1, $\frac{dy}{dx} = \text{increment in } y \text{ multiplied by } 10$. Plotting $x_{r+1/2}$ and $\frac{dy}{dx}$, Plate XI, gives a line whose x -intercept is .3 which is therefore the value of b . The column $x - b$ is then filled out and corresponding values of $\frac{dy}{dx}$ and $x_{r+1/2}$ are plotted on logarithm paper, (Plate XII). Since there is no logarithm of a negative number only absolute value is considered and the sign is taken account of afterwards as always in working with logarithms.



x	y	$\frac{dy}{dx}$	x - b
0	3.20		
.05		.5	-2.5
.1	3.23		
.15		.2	-1.5
.2	3.25		
.25		.1	- .05
.3	3.26		
.35		-.1	.05
.4	3.25		
.45		-.2	.15
.5	3.25		
.55		-.5	.25
.6	3.18		
.65		-.5	.35
.7	3.13		
.75		-.7	.45
.8	3.06		
.85		-.8	.55
.9	2.98		

Plotting corresponding values of $-\frac{dy}{dx}$ and $x_r + 1/2$ on logarithm paper and drawing the best straight line to represent this data gives the y-intercept = $nk = 1.26$ and the slope = $n-1 = .846$

$$n = 1.846$$

and $k = .683$

Taking account now of the fact that $\frac{dy}{dx}$ and $x - b$ had opposite signs we have really determined the constants in

$$a - y = k(x - b)^n$$

and substituting values of k , b and n

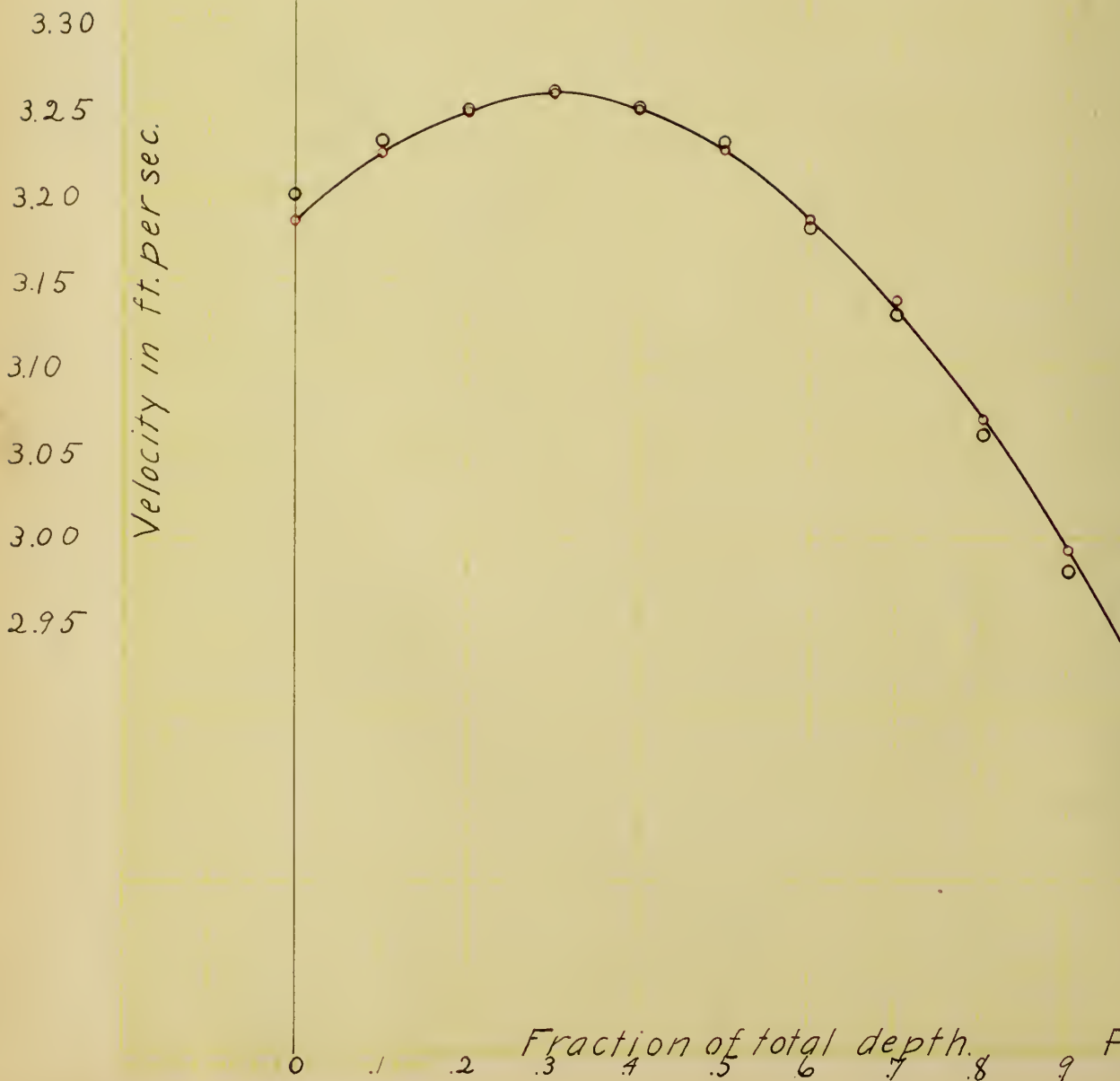
$$a - y = .683(x - .3)^{1.846}$$

Curve showing relation of
Velocity and Depth
in
Mississippi River.

$$Y = 3.259 + .683(x-.3)^{1.846}$$

○ Observed.

● Computed.



where absolute value and not sign of $(x - .3)^{1.846}$ is considered.

Corresponding values of x and y substituted in this equation gave the values of a indicated in column a of the table. The error in the value of y arising from adopting the average value of a , that is the difference between an observed y and one computed by

$$3.259 - y = .683(x - .3)^{1.846}$$

is tabulated in column marked v .

x	y	a	v	% error
0	3.20	3.274	.015	.5
.1	3.23	3.265	.006	.2
.2	3.25	3.260	.001	.0
.3	3.26	3.260	.001	.0
.4	3.25	3.260	.001	.0
.5	3.23	3.265	.006	.2
.6	3.18	3.254	-.005	.2
.7	3.13	3.256	-.003	.1
.8	3.06	3.250	-.009	.3
.9	2.98	<u>3.246</u>	-.013	<u>.4</u>
Average		3.259	Average	.2

Plate XIII

The elimination of a and b as performed in this problem is due to Prof. Langsdorf and has been published by him in The Journal of the Association of Engineering Societies. The

problem has been taken up here as an illustration of the possibilities in the use of logarithm paper. It is especially useful just where other methods of curve-fitting fail, namely in finding of an unknown exponent.

ON THE FORM OF FUNCTION TO BE ASSUMED.

While this paper has so far dealt with the method of fitting a previously assumed function to certain data by a suitable determination of its constants the important question as to the choice of form of function remains unanswered. Concerning the five types of function suitable for frequency curves mentioned on page 13, Prof. Pearson has set up analytical criteria to determine which should be chosen in any particular case. It would be of great importance if analogous tests could be set up for curves suited to engineering and physical data, but little has been done in studying this problem. While nothing approaching a systematic treatment of this whole subject is attempted here, there will be presented certain tests for some of the commoner functions based on the calculus of finite differences.

Let

$$y = f(x)$$

and $(y_0, x_0), (y_1, x_0 + h), (y_2, x_0 + 2h) \dots (y_n, x_0 + nh)$

be corresponding values of x and y taken at equal distances

along x-axis.

Let

$$D'_0 = y_1 - y_0, \quad D'_1 = y_2 - y_1 - - - \quad D'_k = y_{k+1} - y_k - - -$$

$$D^2_0 = D'_1 - D'_0, \quad D^2_1 = D'_2 - D'_1 - - - \quad D^2_k = D'_{k+1} - D'_k - - -$$

Then

$$D^2_0 = y_2 - 2y_1 + y_0$$

$$D^3_0 = D^2_1 - D^2_0$$

$$= y_3 - 3y_2 + 3y_1 - y_0$$

Hence since successive coefficients are formed by same additions as coefficients in powers of a binomial,

$$D^r_0 = y_r - ry_{r-1} + \frac{r(r-1)}{2!} y_{r-2} - - - \text{to } (r+1) \text{ terms.}$$

$$D'_s = D'_{s-1} + D^2_{s-1}$$

$$= D'_{s-2} + 2D^2_{s-2} + D^3_{s-2}$$

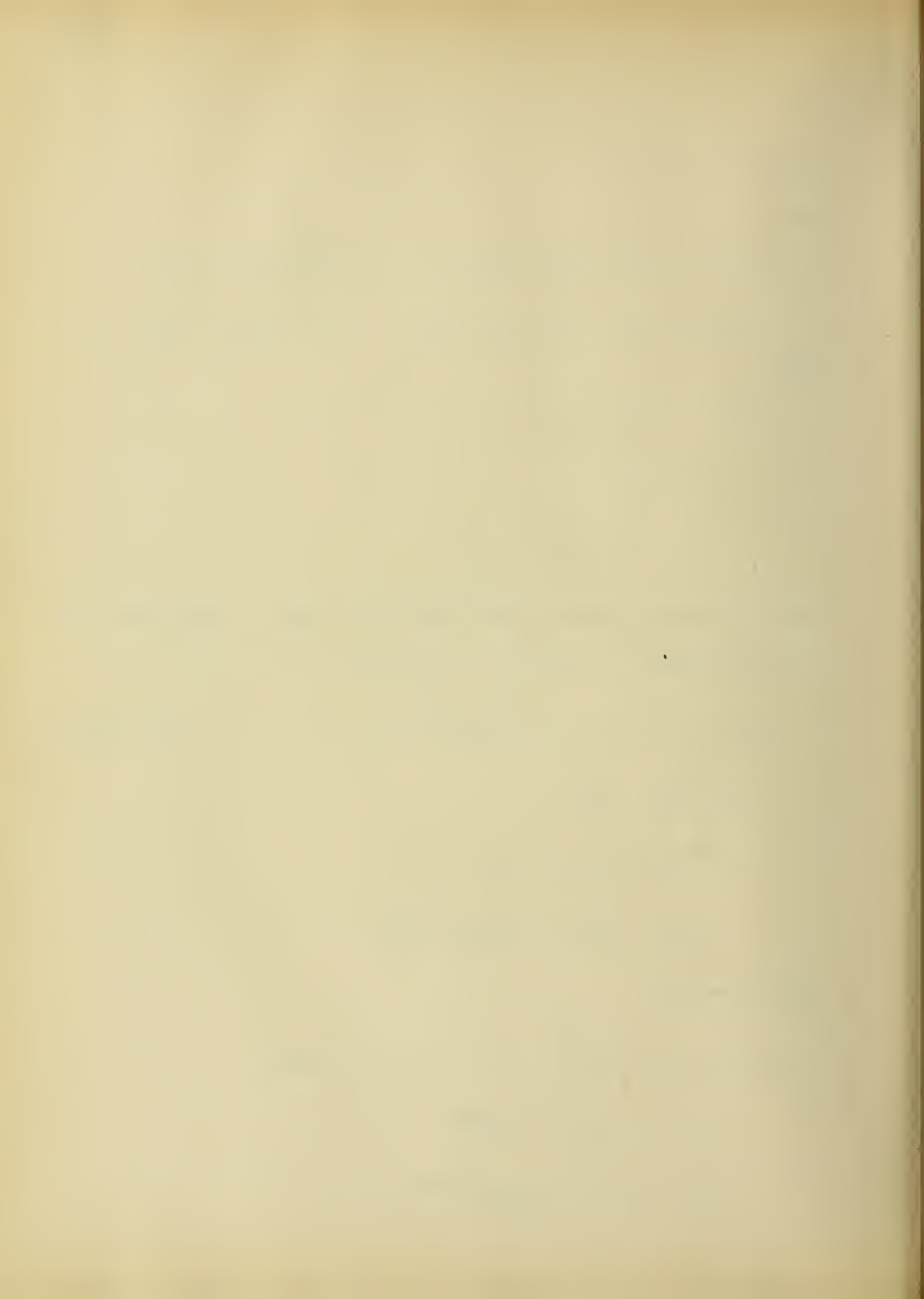
$$= D'_{s-3} + 3D^2_{s-3} + 3D^3_{s-3} + D^4_{s-3}$$

And in general

$$I \quad D^r_s = D^r_{s-1} + D^{r+1}_{s-1} - - - \text{to } (r+1) \text{ terms.}$$

$$= D^r_{s-2} + 2D^{r+1}_{s-2} + D^{r+2}_{s-2} - - - -$$

$$= D^r_0 + sD^{r+1}_0 + \frac{s(s-1)}{2!} D^{r+2}_0 + - - -$$



This discussion will now be applied to a parabolic curve,

$$y = a_0 + a_1x + a_2x^2 + \dots + a_nx^n$$

Since (x_0, y_0) , (x_0+h, y_1) - - - (x_0+nh, y_n) are on the curve

$$y_0 = a_0 + a_1x_0 + a_2x_0^2 + \dots + a_nx_0^n$$

$$y_1 = a_0 + a_1(x_0+h) + a_2(x_0+h)^2 + \dots + a_n(x_0+h)^n$$

- - - - -

$$y_n = a_0 + a_1(x_0+nh) + a_2(x_0+nh)^2 + \dots + a_n(x_0+nh)^n$$

$$D_0^1 = a_1h + a_2(2x_0h+h^2) + \dots + a_n[(x_0+h)^n - x_0^n]$$

$$D_1^1 = a_1h + a_2(2x_0h+3h^2) + \dots + a_n[(x_0+2h)^n - (x_0+h)^n]$$

- - - - -

All first order differences are therefore of degree $(n-1)$ in x and are all independent of a_0 . Moreover the first term in each first order difference is a_1h . The first term in the coefficient of any a as a_2 is $rx_0^{n-1}h$.

Therefore

$$D_0^2 = 2a_2h^2 + \dots$$

of degree $n-2$ in x and



$$D_0^n = n! a_n h^n$$

for each order of differences drops an a from the beginning and one degree in x .

Therefore if a set of data is to be exactly fitted by a curve of the parabolic type of n th degree, the n th order of differences is constant and accordingly higher order of differences vanish.

It will now be shown that this condition here proved necessary is also sufficient. By definition of this notation,

$$y_1 = y_0 + D'_0$$

$$y_2 = y_1 + D'_1$$

$$= y_0 + 2D'_0 + D_0^2$$

$$y_3 = y_2 + D'_2 = y_0 + 3D'_0 + 3D_0^2 + D_0^3$$

Thus the coefficients are apparently those of the binomial theorem. It will be assumed that y_k has such coefficients and proved that y_{k+1} will have.

$$\text{II} \quad y_k = y_0 + kD'_0 + \frac{k(k-1)}{2!} D_0^2 + - - - - + D_0^k$$

$$\text{Then } y_{k+1} = y_k + D'_k$$

From I

$$D'_k = D'_0 + kD_0^2 + \frac{k(k-1)}{2!} D_0^3 + - - -$$

Adding II

$$\begin{aligned}
 y_{k+1} &= y_0 + (1+k)D'_0 + \left(k + \frac{k(k-1)}{2!}\right)D_0^2 \\
 &\quad + \left(\frac{k(k-1)}{2!} + \frac{k(k-1)(k-2)}{3!}\right)D_0^{r-1} + \dots \\
 &\quad + \left[\frac{k(k-1) - \dots - (k-r+2)}{r-1!} + \frac{k(k-1) - \dots - (k-r+3)}{r-2!}\right]D_0^{r-1} \\
 &= y_0 + (k+1)D'_0 + \frac{(k+1)k}{2!}D_0^2 + \frac{(k+1)k(k-1)}{3!}D_0^3 \\
 &\quad + \left[\frac{k(k-1) - \dots - (k-r+3)(k-r+2+r-1)}{r-1!}\right]D_0^{r-1} \\
 &= y_0 + (k+1)D'_0 + \frac{(k+1)}{2!}D_0^2 - \dots \\
 &\quad + \frac{(k+1)k(k-1) - \dots - (k-r+3)}{r-1!}D_0^{r-1} + \dots
 \end{aligned}$$

which is the form of coefficients of binomial expansion of $(k+1)$ st power.

Hence the formula holds for $k+1$ if it does for k . But it does hold for $k=3$, hence for all values of k .

Then

$$y_r = y_0 + rD'_0 + \frac{r(r-1)}{2}D_0^2 + \dots$$

But $x_r = x_0 + rh$

$$r = \frac{x_r - x_0}{h}$$

$$r-1 = \frac{x_r - x_0 - h}{h}, \text{ etc.}$$

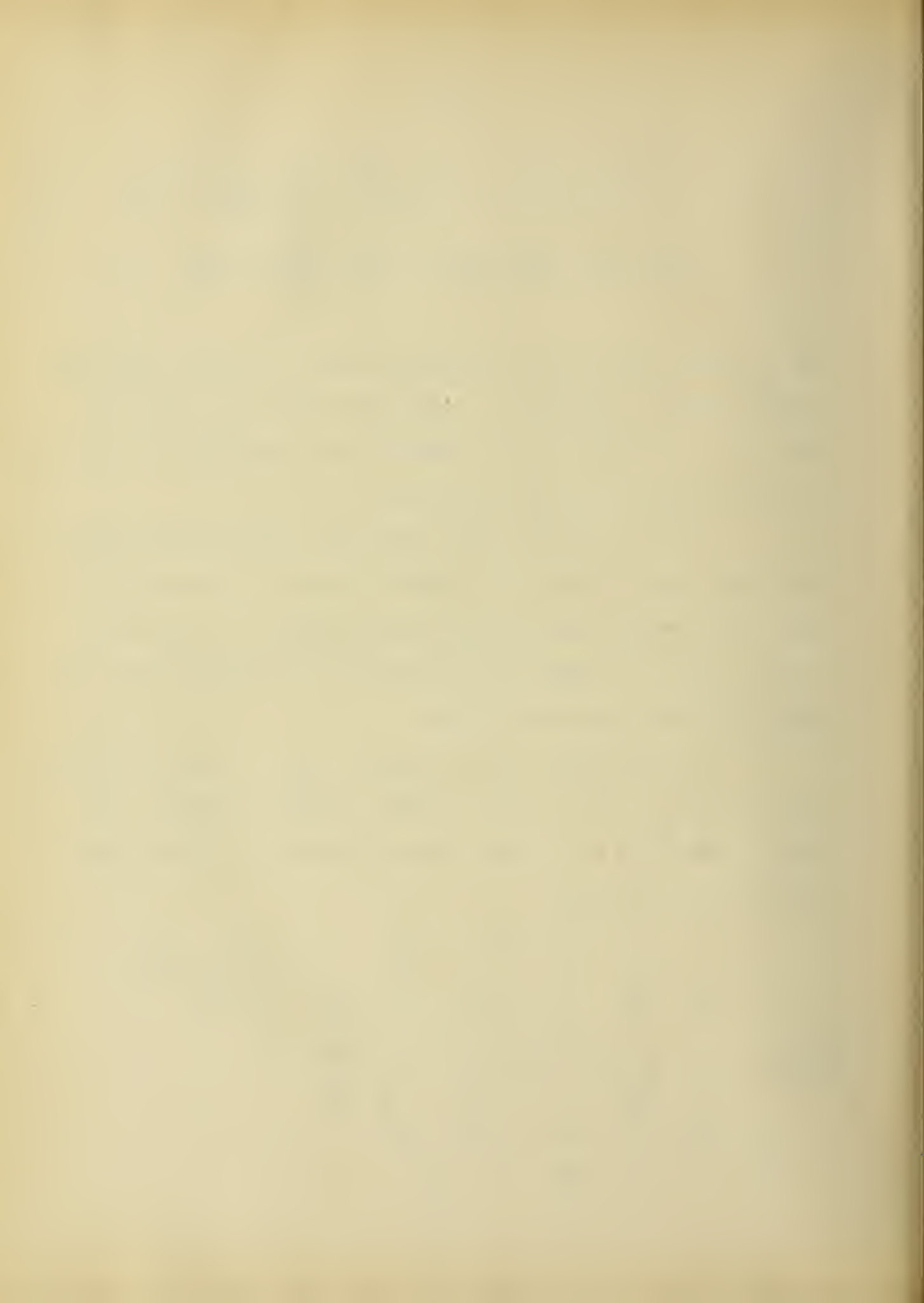
$$y_r = y_0 + \frac{(x_r - x_0)}{h} D'_0 + \frac{(x_r - x_0)}{h} \cdot \frac{(x_r - x_0 - h)}{2h} D''_0 + \frac{(x_r - x_0)}{h} \cdot \frac{(x_r - x_0 - h)}{2h} \cdot \frac{(x_r - x_0 - 2h)}{3h} D'''_0 - \dots$$

But x_r and y_r are any corresponding values of x and y , hence may be the running coordinates. Hence if differences beyond the n th vanish, y is expressed in constants and integral powers of x not extending beyond the n th degree.

Hence a necessary and sufficient condition that certain data may be fitted by a parabolic curve of n th degree is that the $(n+1)$ st order of differences vanish. If the differences do not all vanish but approximate that condition then the curve will be an approximate fit.

To illustrate, the 10 points $(-3,0)$, $(-2,-15)$, $(-1,0)$, $(0,9)$, $(1,0)$, $(2,-15)$, $(3,0)$, $(4,105)$, $(5,384)$, $(6,945)$ can be exactly fitted by a fourth degree parabola, as fifth order of differences vanish.

	-3	-2	-1	0	1	2	3	4	5	6
y	0	-15	0	9	0	-15	0	105	384	945
D'	-15	15	9	-9	-15	15	105	279	561	
D''	30	-6	-18	-6	30	90	174	282		
D'''	-36	-12	12	36	60	84	108			
D''''	24	24	24	24	24	24				
D'''''	0	0	0	0	0					



Successive orders of differences in

$$y = a \sin nx + b \cos nx$$

may be calculated by first getting differences in the simpler functions

$$y = \sin nx \quad \text{and} \quad y = \cos nx$$

x will be given successive increments of unity and y will assume the values

$$\sin(nx), \sin(nx+n), \sin(nx+2n), - - - \sin(nx+kn)$$

$$D'_0 = \sin(nx+n) - \sin(nx)$$

$$= 2 \sin \frac{n}{2} \cos\left(nx + \frac{n}{2}\right)$$

$$= 2 \sin \frac{n}{2} \sin\left(nx + \frac{n+\pi}{2}\right)$$

$$D'_1 = \sin(nx+2n) - \sin(nx+n)$$

$$= 2 \sin \frac{n}{2} \cos\left(nx + \frac{3n}{2}\right)$$

$$= 2 \sin \frac{n}{2} \sin\left(nx + \frac{3n+\pi}{2}\right)$$

$$D'_r = \sin[nx + (r+1)n] - \sin(nx+rn)$$

$$= 2 \sin \frac{n}{2} \cos\left[nx + \left(\frac{2r+1}{2}\right)n\right]$$

$$= 2 \sin \frac{n}{2} \sin\left[nx + \frac{(2r+1)n+\pi}{2}\right]$$



$$\begin{aligned}
D_0^2 &= D_1' - D_0' \\
&= 2 \sin \frac{n}{2} \left[\sin \left(nx + \frac{3n+\pi}{2} \right) - \sin \left(nx + \frac{n+\pi}{2} \right) \right] \\
&= 2 \sin \frac{n}{2} \left[\sin \frac{n}{2} \cos \frac{2n+\pi}{2} \right] \\
&= (2 \sin \frac{n}{2})^2 \sin \left(nx + \frac{2n+2\pi}{2} \right) \\
D_r^2 &= (2 \sin \frac{n}{2})^2 \cos \left[nx + \frac{(2r+2)n+\pi}{2} \right] \\
&= (2 \sin \frac{n}{2})^2 \sin \left[nx + \frac{(2r+2)n+2\pi}{2} \right]
\end{aligned}$$

as is found by taking D_r' from D_{r+1}' .

Similarly by operating on the general term are found

$$\begin{aligned}
D_r^3 &= (2 \sin \frac{n}{2})^3 \sin \left[nx + \frac{(2r+3)n+3\pi}{2} \right] \\
D_r^m &= (2 \sin \frac{n}{2})^3 \sin \left[nx + \frac{(2r+m)n+m\pi}{2} \right]
\end{aligned}$$

For $y = \cos (nx)$

successive values of y are

$\cos (nx), \cos (nx+n)', \cos (nx+2n), - - - \cos (nx+kn)$

$$\begin{aligned}
D_0' &= \cos (nx+n) - \cos nx \\
&= -2 \sin \frac{n}{2} \sin \left(nx + \frac{n}{2} \right) \\
&= 2 \sin \frac{n}{2} \cos \left(nx + \frac{n+\pi}{2} \right)
\end{aligned}$$

$$D_1' = \cos(nx + 2n) - \cos(nx + n)$$

$$= 2 \sin \frac{n}{2} \cos\left(nx + \frac{3n + \pi}{2}\right)$$

$$D_r' = 2 \sin \frac{n}{2} \cos\left[nx + \frac{(2r+1)n + \pi}{2}\right]$$

$$D_r^2 = 2 \sin \frac{n}{2} \left[2 \sin \frac{n}{2} \cos\left(nx + \frac{(2r+2)n + 2\pi}{2}\right) \right]$$

$$= (2 \sin \frac{n}{2})^2 \cos\left[nx + \frac{(2r+2)n + 2\pi}{2}\right]$$

$$D_r^3 = (2 \sin \frac{n}{2})^3 \cos\left[nx + \frac{(2r+3)n + 3\pi}{2}\right]$$

$$D_r^m = (2 \sin \frac{n}{2})^m \cos\left[nx + \frac{(2r+m)n + m\pi}{2}\right]$$

Substituting the values of these successive differences in

$$y = a \sin(nx) + b \cos(nx)$$

$$D_0' = a \left[2 \sin \frac{n}{2} \sin\left(nx + \frac{n + \pi}{2}\right) \right] + b \left[2 \sin \frac{n}{2} \cos\left(nx + \frac{n + \pi}{2}\right) \right]$$

$$= 2 \sin \frac{n}{2} \left[a \sin\left(nx + \frac{n + \pi}{2}\right) + b \cos\left(nx + \frac{n + \pi}{2}\right) \right]$$

$$D_1' = 2 \sin \frac{n}{2} \left[a \sin\left(nx + \frac{3n + \pi}{2}\right) + b \cos\left(nx + \frac{n + \pi}{2}\right) \right]$$

$$D_r' = 2 \sin \frac{n}{2} \left[a \sin\left(nx + \frac{(2r+1)n + \pi}{2}\right) + b \cos\left(nx + \frac{(2r+1)n + \pi}{2}\right) \right]$$

$$D_0^2 = (2 \sin \frac{n}{2})^2 \left[a \sin\left(nx + \frac{2n + 2\pi}{2}\right) + b \cos\left(nx + \frac{2n + 2\pi}{2}\right) \right]$$

$$D_r^2 = (2 \sin \frac{n}{2})^2 \left[a \sin\left(nx + \frac{(2r+2)n + 2\pi}{2}\right) + b \cos\left(nx + \frac{(2r+2)n + 2\pi}{2}\right) \right]$$

$$D_0^m = (2 \sin \frac{n}{2})^m a \left[\sin \left(nx + \frac{m(n+\pi)}{2} \right) + b \cos \left(nx + \frac{m(n+\pi)}{2} \right) \right]$$

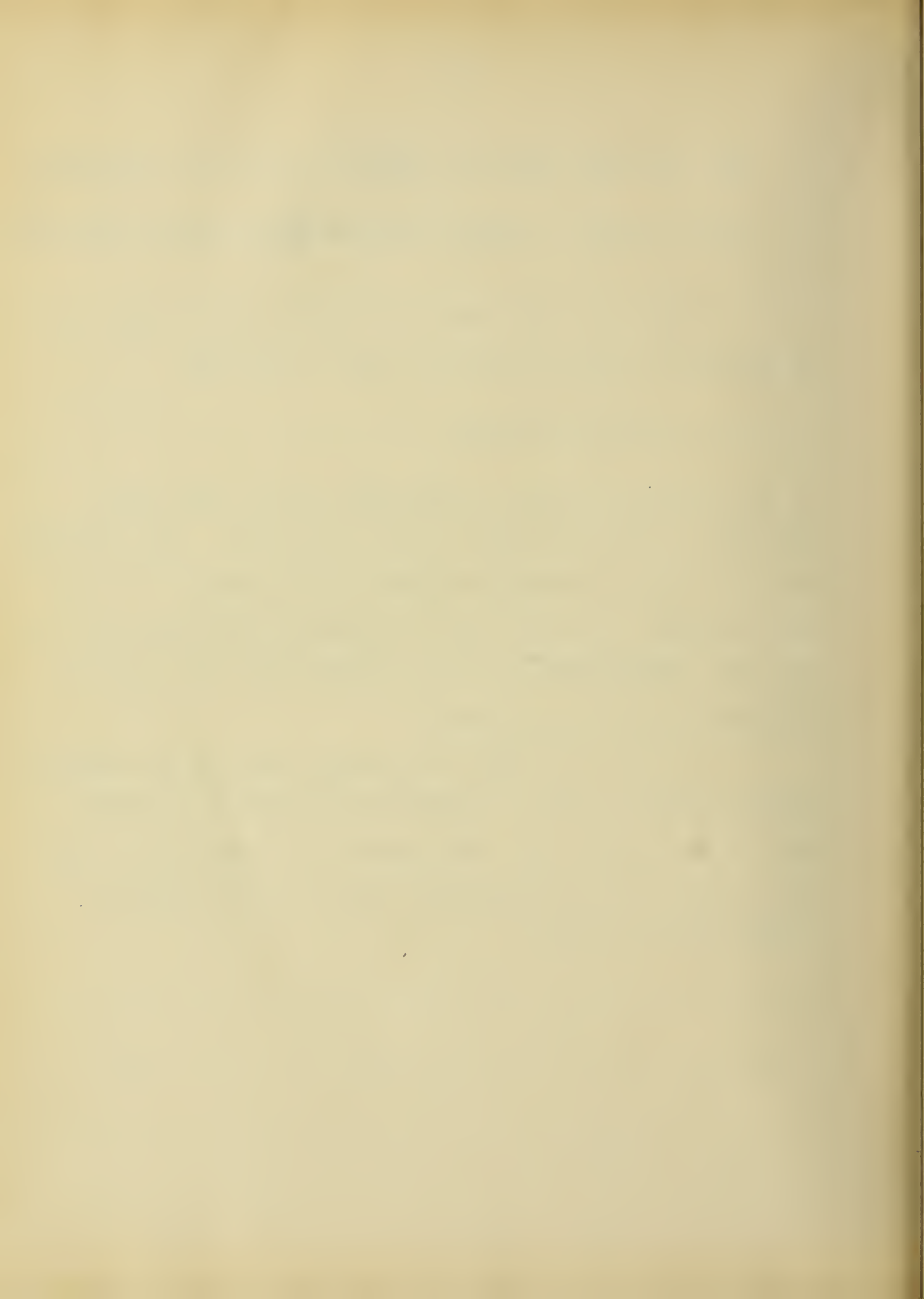
$$D_r^m = (2 \sin \frac{n}{2})^m a \sin \left(nx + \frac{(2r+m)n+m\pi}{2} \right) + b \cos \left(nx + \frac{(2r+m)n+m\pi}{2} \right)$$

The result shows that a necessary condition that a set of data may be exactly fitted by a curve of the form

$$y = a \sin(nx) + b \cos(nx)$$

is that successive orders of differences be also sine curve functions of period the same as the original data. The vertical scale is changed by successive powers of the constant $2 \sin \frac{n}{2}$ and the initial point on x-axis is shifted by the introduction of the term $\frac{(2r+m)n+m\pi}{2}$ but the essential fact of a repeating function of same period persists.

So far as is known to the writer tests for the various forms of function suited to engineering and physical problems have not been worked out. The problem is an important one and the calculus of finite differences seems to be the most promising method of attack.



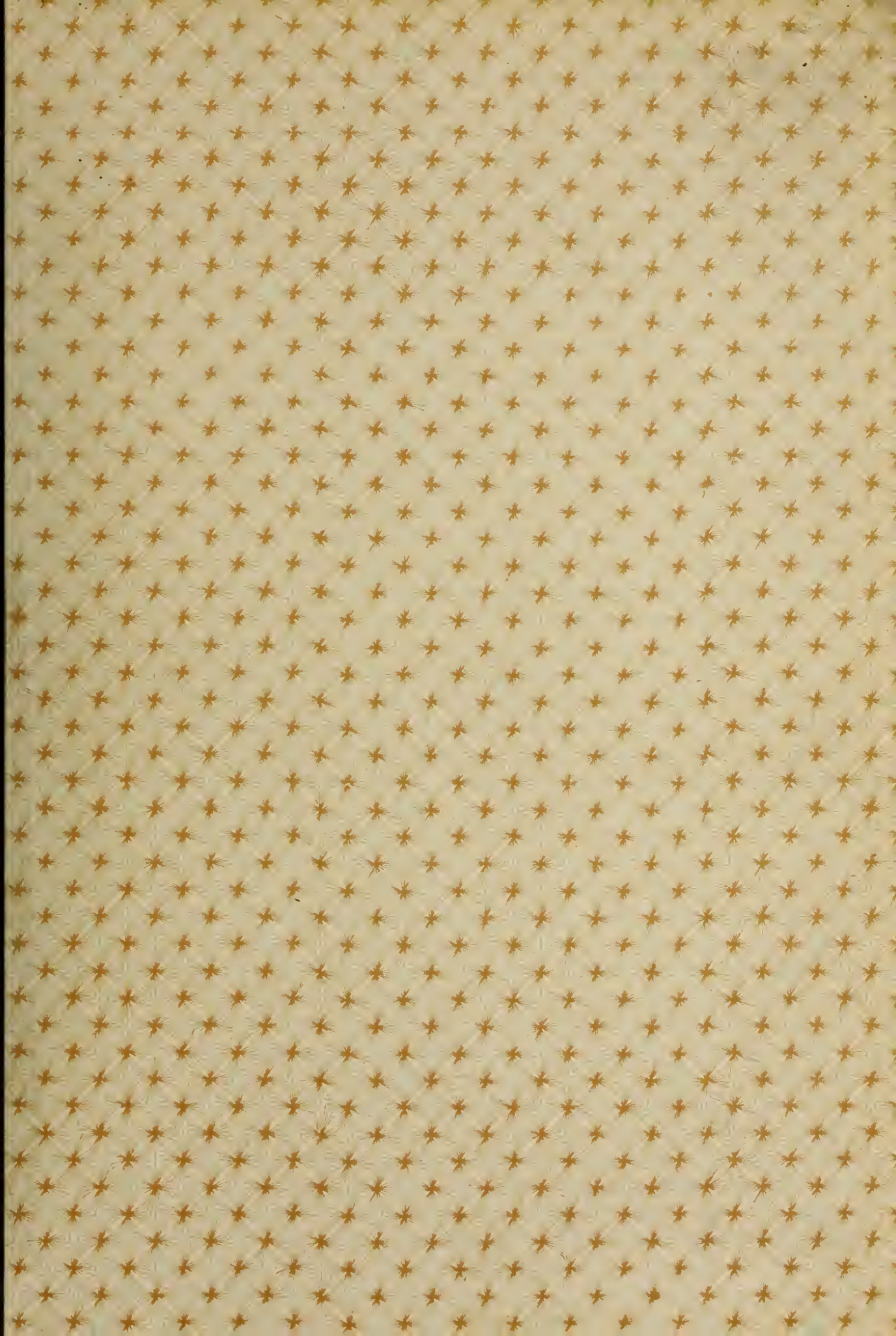
BIBLIOGRAPHY.

-
- Todhunter, I.- History of the Theory of Probability.
- Galton, F.- Inquiries into Human Faculty and its Development.
Natural Inheritance.
Family Likeness in Stature. Proc. of Royal Soc., 1886. 1888.
- Merriman, M.- Method of Least Squares.
- Edgeworth, Prof. F. Y. - In the London, Edinburgh and Dublin Philosophical Review. Vols. 21. 22, 23, 24, 25, 30, 34, 35, 36, 41.
In the Journal of the Royal Statistical Society 1886, 1880, 1890, 1893, 1894, 1895, 1896, 1897, 1898, 1899, 1900.
- Foss, Wm. E.- New Formulas for Calculating the Flow of Water in Pipes and Channels. In Journal of the Association of Engineering Societies, June, 1894, Vol. XIII, No. 6.



- Bowley, A. L. - Elements of Statistics, Part II.
- Pearson, Prof. K.- In Transactions of Royal Society,
1894, 1895, 1896, 1898.
Stat. Soc. Journal, 1896, 1897.
Biometrika, Vol. I, part III, April,
1902, pp. 266-303.
Biometrika, Vol. II, part I, Nov. 1902,
pp. 1-23.
- Boole, Geo.- Calculus of Finite Differences.
- Langsdorf- Curve Fitting in Journal of the Associ-
ation of Engineering Societies, June,
1904.





UNIVERSITY OF ILLINOIS-URBANA



3 0112 086763494